

# The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOEN, JR., *Teachers College, Columbia University*

## CONSULTING EDITORS

GORDON W. ALLPORT

*Harvard University*

URIE BRONFENBRENNER

*Cornell University*

DONALD T. CAMPBELL

*Northwestern University*

LAUNOR F. CARTER

*The RAND Corporation*

RICHARD CHRISTIE

*Columbia University*

MORTON DEUTSCH

*Bell Telephone Laboratories*

I. E. FARBER

*State University of Iowa*

LEON FESTINGER

*Stanford University*

DONALD W. FISKE

*University of Chicago*

ROBERT R. HOLT

*New York University*

HERBERT H. HYMAN

*Columbia University*

IRVING L. JANIS

*Yale University*

HAROLD E. JONES

*University of California*

HAROLD H. KELLEY

*University of Minnesota*

LEONARD S. KOGAN

*Community Service Society  
of New York*

GARDNER LINDZEY

*University of Minnesota*

DAVID C. MCCLELLAND

*Harvard University*

QUINN MCNEMAR

*Stanford University*

LEO POSTMAN

*University of California*

EDMUND H. RODNICK

*Duke University*

DEVITT SANFORD

*University of California*

SEYMOUR SARASON

*Yale University*

RICHARD L. SOLOMON

*Harvard University*

RENATO TAGIURI

*Harvard University*

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: BARBARA CUMMINGS; FRANCES H. CLARK; SADIE J. DOYLE, SARAH WOMACK

VOLUME 55  
1957

PUBLISHED BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.  
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1957 by the American Psychological Association, Inc.



# CONTENTS OF VOLUME 55

## ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

	PAGE
ALEXANDER, I. E., & BLACKMAN, S. Castration, Circumcision, and Anti-Semitism.....	143
ALLPORT, G. W., & PETTIGREW, T. F. Cultural Influence on the Perception of Movement: The Trapezoidal Illusion Among Zulus.....	104
ALPER, THELMA G. Predicting the Direction of Selective Recall: Its Relation to Ego Strength and Achievement.....	149
APPLEZWEIG, M. H. See MOELLER, G.	
ARKOFF, A. Resolution of Approach-Approach and Avoidance-Avoidance Conflicts.....	402
ARNHOFF, F. N. Ethnocentrism and Stimulus Generalization: A Replication and Further Study..	393
BACON, MARGARET K. See BARRY, H., III.	
BARCH, A. M., TRUMBO, DON, & NANGLE, J. Social Setting and Conformity to a Legal Requirement..	396
BARRY, H., III, BACON, MARGARET K., & CHILD, I. L. A Cross-Cultural Survey of Some Sex Differences in Socialization.....	327
BERKOWITZ, L. Effects of Perceived Dependence Relationships Upon Conformity to Group Expectations.....	350
BERKOWITZ, L. See LUNDY, R. M.	
BINDER, A., MCCONNELL, D., & SJOHOLM, NANCY A. Verbal Conditioning as a Function of Experimenter Characteristics.....	309
BINDON, D. MARJORIE. Make-A-Picture Story (MAPS) Test Findings for Rubella Deaf Children....	38
BLACKMAN, S. See ALEXANDER, I. E.	
BLOCK, J. A Study of Affective Responsiveness in a Lie-Detection Situation.....	11
BORN, J. J. See SIDMAN, M.	
BOSSOM, J., & MASLOW, A. H. Security of Judges as a Factor in Impressions of Warmth in Others....	147
BROWN, R. W. Linguistic Determinism and the Part of Speech.....	1
BRUNER, J. S., & PERLMUTTER, H. V. Compatriot and Foreigner: A Study of Impression Formation in Three Countries.....	253
CARON, A. J., & WALLACH, M. A. Recall of Interrupted Tasks Under Stress: A Phenomenon of Memory or of Learning?.....	372
CHILD, I. L. See BARRY, H., III.	
CLEVELAND, S. See WARE, K. E.	
COLLIER, R. M. Consciousness as a Regulatory Field: A Theory of Psychotherapy.....	275
CROW, W. J. The Effect of Training Upon Accuracy and Variability in Interpersonal Perception....	355
DAVIS, R. H. A Further Study of the Effect of Stress on Palmar Prints.....	132
DECHARMS, R. Affiliation Motivation and Productivity in Small Groups.....	222
DENNIS, W. Uses of Common Objects as Indicators of Cultural Orientations.....	21
DITTES, J. E. Galvanic Skin Response as a Measure of Patient's Reaction to Therapist's Permissiveness.....	295
DAULNY, D. E., JR. Avoidance Learning of Perceptual Defense and Vigilance.....	333
EDWARDS, A. L. Social Desirability and Probability of Endorsement of Items in the Interpersonal Check List.....	394
ERRATA.....	273
EXLINE, R. V. Group Climate as a Factor in the Relevance and Accuracy of Social Perception....	382
FESHBACH, S., & SINGER, R. D. The Effects of Fear Arousal and Suppression of Fear Upon Social Perception.....	283
FISHER, S. See WARE, K. E.	
FUNKENSTEIN, D. H. See KING, S. H.	
GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The Psychological Meaning of Acquiescence Set for Authoritarianism.....	98
GRICE, G. R. On the Relations Between Discrimination Reaction Time, Anxiety, and Intelligence..	147
GRISELL, J. L. See ROSENBAUM, G.	
HARE, A. P. Situational Differences in Leader Behavior.....	132
HARRIS, D. B. A Scale for Measuring Attitudes of Social Responsibility in Children.....	322
HART, I. Maternal Child-Rearing Practices and Authoritarian Ideology.....	232
HARVEY, O. J. See HOVLAND, C. I.	
HARWOOD, A. See MEDNICK, S. A.	
HILL, W. F. See SCHEIN, E. H.	
HORST, P. See STROTHER, C. R.	
HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and Contrast Effects in Reactions to Communication and Attitude Change.....	244
KAMIYA, J. See STONE, P.	
KAPLAN, B. On the Phenomena of "Opposite Speech".....	389
KELLEY, H. H., & LAMB, T. W. Certainty of Judgment and Resistance to Social Influence.....	137



KING, S. H., & FUNKENSTEIN, D. H. Religious Practice and Cardiovascular Reactions During Stress.	135
LAMB, T. W. See KELLEY, H. H.	
LANZETTA, J. T., & ROBY, T. B. Group Learning and Communication as a Function of Task and Structure "Demands".	121
LEAVITT, G. S. See GAGE, N. L.	
LESSER, G. S. The Relationship Between Overt and Fantasy Aggression as a Function of Maternal Response to Aggression.	218
LEVENTHAL, D. See MORAN, L. J.	
LEVENTHAL, H. Cognitive Processes and Interpersonal Predictions.	176
LEVIN, H., & TURGEON, VALERIE F. The Influence of the Mother's Presence on Children's Doll Play Aggression.	304
LIVSON, N., & MUSSEN, P. H. The Relation of Ego Control to Overt Aggression and Dependency.	66
LUBIN, A. See SCHEIN, E. H.	
LUNDY, R. M., & BERKOWITZ, L. Cognitive Complexity and Assimilative Projection in Attitude Change.	34
LYKKEN, D. T. A Study of Anxiety in the Sociopathic Personality.	6
MACCOBY, ELEANOR E., & WILSON, W. C. Identification and Observational Learning from Films.	76
MACHI, V. S. See WILLIAMS, R. J.	
MACKAVEY, W. R. See ROSENBAUM, G.	
MANDLER, G., & PARNES, ELIZABETH W. Frequency and Idiosyncrasy of Associative Responses.	58
MARTIN, B. See TRUAX, C. B.	
MASLOW, A. H. See BOSSOM, J.	
MCCONNELL, D. See BINDER, A.	
MCDIARMID, C. G. See STEINER, I. D.	
MCGAUGHRAN, L. S. See MORAN, L. J.	
MEDNICK, MARTHA T. Mediated Generalization and the Incubation Effect as a Function of Manifest Anxiety.	315
MEDNICK, S. A., HARWOOD, A., & WERTHEIM, J. Perception of Disturbing and Neutral Words Through the Autokinetic Word Technique.	267
MENZEL, H. Public and Private Conformity Under Different Conditions of Acceptance in the Group.	398
MEYER, V. The Treatment of Two Phobic Patients on the Basis of Learning Principles.	261
MILLON, T. Authoritarianism, Intolerance of Ambiguity, and Rigidity Under Ego- and Task-Involving Conditions.	29
MILTON, G. A. The Effects of Sex-Role Identification Upon Problem-Solving Skill.	208
MOELLER, G., & APPLEZWEIG, M. H. A Motivational Factor in Conformity.	114
MORAN, L. J., MCGAUGHRAN, L. S., & LEVENTHAL, D. Intra-Individual Consistency in Conceptualization.	194
MUSSEN, P. H. See LIVSON, N.	
NANGLE, J. See BARCH, A. M.	
PARNES, ELIZABETH W. See MANDLER, G.	
PERLMUTTER, H. V. See BRUNER, J. S.	
PETTIGREW, T. F., See ALLPORT, G. W.	
PRENTICE, N. M. The Influence of Ethnic Groups.	270
REED, M. R. The Masculinity-Femininity Dimension in Normal and Psychotic Subjects.	289
ROBY, T. B. See LANZETTA, J. T.	
ROSENBAUM, G., GRISSELL, J. L., & MACKAVEY, W. R. The Relationship of Age and Privilege Status to Reaction Time Indices of Schizophrenic Motivation.	202
ROSNER, S. Consistency in Response to Group Pressures.	146
SALZINGER, K. Shift in Judgment of Weights as a Function of Anchoring Stimuli and Instructions in Early Schizophrenics and Normals.	43
SAMELSON, F. Conforming Behavior Under Two Conditions of Conflict in the Cognitive Field.	181
SCHAE, K. W. See STROTHER, C. R.	
SCHEIN, E. H., HILL, W. F., WILLIAMS, H. L., & LUBIN, A. Distinguishing Characteristics of Collaborators and Resisters Among American Prisoners of War.	197
SCOTT, W. A. Attitude Change Through Reward of Verbal Behavior.	72
SHERIF, M. See HOVLAND, C. I.	
SIDMAN, M., & BOREN, J. J. The Relative Aversiveness of Warning Signal and Shock in an Avoidance Situation.	339
SIEGEL, ALBERTA E., & SIEGEL, S. Reference Groups, Membership Groups, and Attitude Change.	360
SIEGEL, S. See SIEGEL, ALBERTA E.	
SILLER, J. Socioeconomic Status and Conceptual Thinking.	365
SILVERMAN, R. E. The Manifest Anxiety Scale as a Measure of Drive.	94
SINGER, R. D. See FESHBACH, S.	
SJOHOLM, NANCY A. See BINDER, A.	

SMITH, E. E. The Effects of Clear and Unclear Role Expectations on Group Productivity and Defensiveness.....	213
STAATS, A. W. Learning Theory and "Opposite Speech".....	268
STEINER, I. D., & McDIARMID, C. G. Two Kinds of Assumed Similarity Between Opposites.....	140
STONE, G. C. See GAGE, N. L.	
STONE, PHIL, & KAMIYA, J. Judgments of Consensus During Group Discussion.....	171
STROTHER, C. R., SCHAIK, K. W., & HORST, P. The Relationship Between Advanced Age and Mental Abilities.....	166
TAMKIN, A. S. Selective Recall in Schizophrenia and Its Relation to Ego Strength.....	345
TANNENBAUM, A. S. Personality Change as a Result of an Experimental Change of Environmental Conditions.....	404
TRUAX, C. B. The Repression Response to Implied Failure as a Function of the Hysteria-Psychasthenia Index.....	188
TRAUX, C. B., & MARTIN, B. The Immediate and Delayed Effect of Failure as a Function of Task Complexity and Personalization of Failure.....	16
TRUMBO, DON. See BARCH, A. M.	
TURGEON, VALERIE F. See LEVIN, H.	
WALLACH, M. A. See CARON, A. J.	
WARE, K. E., FISHER, S., & CLEVELAND, S. Body-Image Boundaries and Adjustment to Poliomyelitis.....	88
WERTHEIM, J. See MEDNICK, S. A.	
WILLIAMS, H. L. See SCHEIN, E. H.	
WILLIAMS, R. J., & MACHI, V. S. An Analysis of Interperson Correlations Among Thirty Psychotics.....	50
WILSON, W. C. See MACCOBY, ELEANOR E.	
WORCHEL, P. Catharsis and the Relief of Hostility.....	238
WYLIE, RUTH C. Cognitive Set and Motivational Factors in the Perception of Neutral and Threat-Related Stimuli.....	227



# ARE THERE GAPS IN YOUR FILES OF APA JOURNALS

---

## THEN HEAR THIS . . .

The American Psychological Association announces a sale during the period October 1957 through March 1958. Of the following journals, all *available* issues in the volumes for the years preceding 1951 will be offered at a price of only 50¢ (foreign, 60¢) per issue:

*American Psychologist*  
*Journal of Abnormal & Social Psychology*  
*Journal of Applied Psychology*  
*Journal of Comparative & Physiological Psychology* (1947-1950 only)  
*Journal of Consulting Psychology*  
*Journal of Experimental Psychology*  
*Psychological Abstracts*  
*Psychological Bulletin*  
*Psychological Index* (a few complete volumes, some shopworn)  
*Psychological Monographs*  
*Psychological Review*

---

Not *all* issues in *all* volumes are available. But—ORDER NOW before more back issues go out of print. From our available stock we will complete as much of your order as possible at this reduced price and for this limited period.

Delivery: 6 to 8 weeks

No dealer or quantity discounts

After this sale, for the years preceding 1948, journals will be available only on microfilm and microcard.

Order from:

American Psychological Association  
Department BB, 1333 Sixteenth Street, N.W.  
Washington 6, D. C.

S  
A  
L  
E  
  
£  
  
B  
A  
C  
K  
  
I  
S  
S  
U  
E  
S

# The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOEN, JR., *Teachers College, Columbia University*

## CONSULTING EDITORS

GORDON W. ALLPORT  
*Harvard University*

URIE BRONFENBRENNER  
*Cornell University*

DONALD T. CAMPBELL  
*Northwestern University*

LAUNOR F. CARTER  
*The RAND Corporation*

RICHARD CHRISTIE  
*Columbia University*

MORTON DEUTSCH  
*Bell Telephone Laboratories*

I. E. FARBER  
*State University of Iowa*

LEON FESTINGER  
*Stanford University*

DONALD W. FISKE  
*University of Chicago*

ROBERT R. HOLT  
*New York University*

HERBERT H. HYMAN  
*Columbia University*

IRVING L. JANIS  
*Yale University*

HAROLD E. JONES  
*University of California*

HAROLD H. KELLEY  
*Center for Advanced Study  
in the Behavioral Sciences*

LEONARD S. KOGAN  
*Community Service Society  
of New York*

GARDNER LINDZEY  
*Syracuse University*

DAVID C. MCCLELLAND  
*Harvard University*

QUINN MCNEMAR  
*Stanford University*

LEO POSTMAN  
*University of California*

ELIOT H. RODNICK  
*Duke University*

NEVITT SANFORD  
*University of California*

SEYMOUR SARASON  
*Yale University*

RICHARD L. SOLOMON  
*Harvard University*

RENATO TAGIURI  
*Harvard University*

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: BARBARA CUMMINGS, FRANCES H. CLARK, SADIE J. DOYLE

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

Bureau Ednl. Psy. Research  
DAVID HALL TRAINING COLLEGE  
Dated .....  
Acas. No .....



The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is \$16.00; foreign, \$16.50. Single issues are \$3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH  
Graduate Department of Psychology  
New York University  
Washington Square  
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 Sixteenth Street N.W.  
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.  
Baltimore 2, Maryland

PUBLISHED BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
Mt. Royal and Guilford Aves., Baltimore 2, Md.  
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1957 by the American Psychological Association, Inc.

# LINGUISTIC DETERMINISM AND THE PART OF SPEECH<sup>1</sup>

ROGER W. BROWN

*Harvard University*

**I**N RECENT years the anthropologists Whorf (12), Lee (9), and Hoiyer (6) have put forward the view that language is a determinant of perception and thought. The nature of the determining influence exerted by the vocabulary of a language is quite clear (1), but it is less easy to see how the grammatical features of a language can affect cognition. Yet it is just the grammatical differences between languages that are most striking and it is their determining force that the anthropologist has stressed. This paper undertakes to show how one kind of grammatical practice, the allocation of words to one or another part of speech, does affect cognition.

The words of a language can be collected into classes of formal equivalents which are called the parts of speech. Fries (5) has shown that English nouns, for example, are words acceptable in sentence frames of the type "(The) ——— is (are) good." Native speakers of English will find it possible to insert "concert" or "cow" or "truth" in that frame but will find "very" or "of" grammatically impossible. Fries has used other sentence frames to separate out verbs, adjectives, and adverbs. In French, nouns can be further subdivided into formal gender classes. In Navaho there are more than 20 formal classes for words naming different kinds of objects. All of these word classes are defined by linguistic science in terms of the combinational possibilities of forms in a language without reference to the meanings of forms.

So long as these classes are defined in purely formal terms they do not suggest important cognitive differences. That suggestion comes in when we add the semantic correlates of the classes. The native speaker of English is likely to think of the parts of speech in semantic terms. Nouns name substances; verbs name processes; and adjectives name qualities. The genders of certain European languages are usually called masculine, feminine, and neuter,

and these are semantic characterizations. The object classes of Navaho are usually described as words naming round objects, words naming long, thin objects, words naming granular substances, etc. The linguistic determinists in anthropology believe that the semantic character of the form classes fixes the fundamental conception of reality in a language community and that differences on this level correspond to different *Weltanschauungen*.

At the same time the science of descriptive linguistics refuses to define its word classes in semantic terms. Fries (5) has shown that for the English parts of speech such definitions are always either unclear or overextended. We all know the English teacher's characterization of the noun as the name of a person, place, or thing. The terms "person" and "place" are reasonably clear, but do not apply to such nouns as "truth," "odor," and "thought." The meaning of "thing" is so unclear that we cannot tell whether or not odor and thought are things. No one has been able to provide clear semantic definitions that will serve to distinguish every English noun from every verb, adjective, and adverb. It is well known that the "masculine" and "feminine" genders in the European languages include names for objects having no sex. In Navaho, too, the object classes do not show perfect semantic consistency. In short, the semantic definitions of the form classes ignore many exceptions and are unsuitable for the purposes of linguistic science.

When the linguistic scientist sets up his descriptive categories he quite naturally looks for attributes of exceptionless validity, and there are not such semantic attributes for the English parts of speech. However, the layman may operate, in this area as in so many others, with conceptions that take account of probabilistic as well as certain associations. It may be that nouns *tend to have* a different semantic from verbs, and that the native speaker detects this tendency while he is in the process of learning the language. To answer these questions examination was made of the nouns and verbs of young children learning English to see whether there was a semantic distinction be-

<sup>1</sup> The gist of this paper was presented in a conference on "Linguistic meaning" sponsored by the Social Science Research Council and held at Yale University, May 17-18, 1956. The Laboratory of Social Relations, Harvard University, financed the research.



tween the two parts of speech. The distinction proved to be much clearer than it is in the vocabulary of English-speaking adults. The second step was an experiment to find out whether the children were aware of the semantic distinction between nouns and verbs and whether they made any use of the distinction.

#### THE NOUN AND VERB IN CHILDREN'S SPEECH

Harvard pre-school sessions were visited for about a month. There were eight children in each class; two of the classes were limited to children between four and five years while a third class accepted those between three and four. As an observer, the author sat on the side-lines and let the pre-school life swirl about him, recording verbatim all the conversation he could hear. From these records, he made vocabulary lists classified into the parts of speech. It was his impression, on examining this vocabulary, that the nouns and verbs of children were more nearly consistent with the classroom semantic definitions than are the nouns and verbs of adults. Nouns commonly heard were "truck," "blocks," and "teacher." There were no uses of "thought" or "virtue" or "attitude." These observations suggested that as the form classes grow larger they decline in semantic consistency. Perhaps children develop firm, and temporarily reliable, notions about the semantics of nouns and verbs. These notions may stay with them as adults even though they retain only a probabilistic truth.

To compare the character of adult and child vocabularies, the first thousand most frequent words from the Thorndike-Lorge (11) list of adult usage were examined, and also the first thousand most frequent words from the Rinsland (10) list of the vocabulary of children in the first grade. The Rinsland list is based on 4,630 pages of conversation, plus more than a thousand letters and stories. The Rinsland list is much the same as lists compiled independently by the Child Study Committee of the Kindergarten Union (8) and by Horn (7).

The first set of contrasts deals with two reduced lists; nouns found among the first thousand for adults but not for children, compared with nouns among the first thousand for children but not for adults. The set of nouns having clearest "thing" character would seem to be those that are called "concrete" and it is a commonplace to describe the language of

children as more concrete than that of adults. One sense of the pair "concrete-abstract" is the same as "subordinate-superordinate." The more abstract term, the superordinate, includes in its denotation the denotation of the concrete or subordinate term, but extends beyond it. Superordinate-subordinate relations between the two lists were all in one direction. The adult list included "action," "article," "body," "experience," and at least seven others which were superordinate to many words on the children's list. There were no nouns on the children's list superordinate to those on the adult list.

The concrete noun with the smaller denotation is likely to be more picturable than its superordinate, and picturability is another common sense of "concrete." Of course the concrete noun, like the abstract, names a category rather than a particular instance. However, some categories have a more or less characteristic visual contour and size while others do not. Visual contour is a defining attribute for "table," but not for "thing" or "experience." Of the adult nouns, 16% named categories having a characteristic visual contour, while 67% of the children's nouns were of this kind. Nouns like "apple," "barn," and "airplane" name categories for which size is a defining attribute, while nouns like "affair," "amount," and "action" do not. On the adult list, 39% of the nouns were of the former kind, while 83% of the children's nouns had size implications. It appears that children's nouns are more likely to name concrete things (in the sense of naming narrow categories with characteristic visual contour and size) than are the nouns of adults.

Two lists of verbs were compared: those among the first 1,000 for adults but not for children, and those among the first 1,000 for children but not for adults. The question here was the percentage of verbs naming animal (including human) movement. Of the adult verbs, 33% were of this kind, while 67% of the children's verbs named actions. The common notion that verbs name actions seems to be truer for the vocabulary of children than for the vocabulary of adults.

These studies of word lists confirm the impression that the nouns and verbs used by children have more consistent semantic implications than those used by adults. It remains



a question whether children are, in any sense, aware of these implications. There are many ways in which such awareness could be useful to one learning the language. Adults often try to convey the sense of a word by speaking it in the presence of the object or event named. All such single namings are ambiguous. The adult who says "water" while looking at a glass of water may cause a child to attend to the glass itself as a container, to the glass as a transparent material, to the liquid character of its contents, to the height of the liquid, to the state of containment, and so on. Selection of the nonlinguistic attributes that govern proper denotative use of the word "water" cannot be guaranteed by a single naming. Repeated pointings can, of course, establish the invariant circumstances governing use of the word. If there were nothing to suggest to the child the probably relevant features of the nonlinguistic world, discovery of linguistic meanings would be a very laborious affair. However, a new word is ordinarily introduced in a way that makes its part-of-speech membership clear: "Look at the *dog*" or "See him *running*." If a part of speech has reliable semantic implications it could call attention to the kind of attribute likely to belong to the meaning of the word. A child who had absorbed the semantics of the noun and verb would know, the first time he heard the word "dog," that it was likely to refer to an object having characteristic size and shape, whereas "running" would be likely to name some animal motion. The part-of-speech membership of the new word could operate as a filter selecting for attention probably relevant features of the nonlinguistic world. It seemed that one could learn whether children experience any such filtering of attributes by introducing to them newly invented words assigned to one or another part of speech, and then inquiring about the meanings the words appeared to have.

In the children's speech that had been recorded, nouns and verbs were given proper grammatical treatment. In addition, the children made correct use of a subclass of nouns—the mass nouns. These are words like "dirt," "snow," "milk," and "rice" which are given different grammatical treatment from such particular nouns as "barn," "house," and "dog." For example, when "some" is used with "barn" the noun is in the plural—"some

barns," whereas a mass noun would be in the singular—"some rice." The semantic difference between these two classes of noun is suggested by the designations "mass" and "particular." Mass nouns usually name extended substances having no characteristic size or shape, while particular nouns name objects having size and shape. Many nouns can function in either a mass or particular way with attendant shifts in the speaker's view of the referent. "Some cake" is a chunk of a mass while "some cakes" are either cupcakes or layer cakes arranged in a row. Many words in the vocabulary of psychology have this double potentiality. Although the personologist deplors such usage, the layman speaks of someone having "a lot of personality" or "very little temperament." The professional insists that personality is not an undifferentiated substance of which one can have more or less. Personalities are like cupcakes—all of a size and one to a customer—with only their frostings to make them unique.

In the speech of the pre-school children "milk," and "orange juice," and "dirt" were the most common mass nouns. These were always given correct grammatical treatment. No one said "a milk" or "some dirt." It was decided to work with three functional classes: the particular noun, the mass noun, and the verb.

#### METHOD

The experiment involved three sets of four pictures each.<sup>2</sup> One of these sets will be described in detail. The first picture in the set shows a pair of hands performing a kneading sort of motion, with a mass of red confetti-like material which is piled into and overflowing a blue-and-white striped container that is round and low in shape. The important features of the picture are the kneading action, the red mass, and the blue-and-white round container. The motion would ordinarily be named with a verb (like "kneading"), the mass with a mass noun (like "confetti"), and the container with a particular noun (like "container"). It was assumed that children would have no readily available names for any of these conceptions. Each of the remaining three pictures of this set exactly reproduced one of the three salient features of the first picture, either the motion, the mass, or the container. In order to represent the motion a second time it was necessary to show also a mass and a container. However, the mass was here painted yellow so as not to duplicate the original, and the container was of a different size, shape, and color from the original. The other two sets of pictures involved different content, but always an action, a mass

<sup>2</sup> The author is grateful to Dr. Susan Ervin for painting the pictures and for suggesting colors and forms that would please children.



substance, and a particular object. In one case, the first picture showed hands cutting a mass of cloth with a strange tool. In the third set, hands were shown emptying an odd container of a slushy sort of material.

In overview, the following use was to be made of the three sets of pictures. Children were to be shown the first picture in conjunction with a new word identifiable either as a verb, a mass noun, or a particular noun. Then they would be shown the remaining three pictures of the set and asked to point out the one that pictured again what had been named in the first picture. It was anticipated that when the new word was a verb they would point to the picture of motion, when it was a particular noun they would point to the container, and when it was a mass noun they would point to the extended substance.

Three word stems were used: "niss," "sib," and "latt." If the stem was to function as a verb X would begin by asking: "Do you know what it means to sib?" (Children do not always answer "no" as they ought.) "In this picture" (first picture of a set) "you can see sibbing. Now show me another picture of sibbing" (presenting the other three pictures of the set). If the stem was to function as a particular noun, X began: "Do you know what a sib is?" and proceeded in consistent fashion. If the word was to function as a mass noun, X began: "Have you ever seen any sib?" and went on accordingly.

Each child saw all three sets of pictures and heard each of the word stems; one of them as a particular noun, one as a mass noun, and one as a verb. The combinations of word stem, part-of-speech membership, picture set, and order of presentation were all randomly varied. There were 16 children in all, half of them between three and four years, and half between four and five. They were all acquainted with X by the time the experiment was performed. The procedure was very like the familiar business of looking at a picture book and naming the things seen and was accepted by the children as a kind of game. The game was always played with one child at a time.

## RESULTS AND DISCUSSION

When a new word was introduced as a verb, 10 of the 16 children picked out the picture of movement. When the word was a particular noun, 11 of 16 selected the picture of an object; and when the word was a mass noun, 12 of 16 selected the extended substance. Of the 15 responses that were not correct, four were

TABLE 1  
PICTURE SELECTIONS FOR WORDS BELONGING TO  
VARIOUS PARTS OF SPEECH

Category Depicted	Verbs	Particular Nouns	Mass Nouns
Actions	10	1	0
Objects	4	11	3
Substances	1	2	12
No response	1	2	1

simply failures to answer because of some distraction from the task. The results are summarized in Table 1. A simple test was made to determine the significance of the differences in the pictures selected by children when the new word was a verb, when the word was a particular noun, and when it was a mass noun. For example, the selections made when the word was a verb were dichotomized into pictures of actions and all others. These frequencies for verbs were compared with like frequencies for the choices made when the words were either particular or mass nouns. In other words, the test was to determine whether action pictures were more likely to be selected as referents for new words introduced as verbs than for new words introduced as nouns. Comparable tests were made to see whether particular nouns were associated with pictures of objects and mass nouns with substances. All three of the resultant  $2 \times 2$  tables yielded differences beyond the .005 level of significance when the Fisher-Yates test was applied.

It is well known that children will sometimes do what an adult wishes in a task of this kind though they do not understand the task as the adult does. Consequently, the qualitative results may be more persuasive than the quantitative. In the first trial with the first child, for instance, X showed the picture of cloth being cut by an odd tool and said that there was a "sib" in the picture. Then went on with: "Can you show me another sib?" and while X still fumbled with the other three pictures, his subject swung around and pointed to the steam valve on the end of the radiator saying, "There's a sib." The pictured tool looked very like the steam valve. In another case, X showed the picture of confetti-kneading and said, "There is some latt in this picture," whereupon his subject said: "The latt is spilling." And it was.

Recent experiments with phonetic symbolism (2) and metaphor (3) indicate that semantic rules are not always arbitrary. A word can suggest its meaning because the sound is an echo of the sense or because the word had a prior meaning which is related to the new meaning. The present study suggests that most words have an additional kind of "appropriateness" stemming from their grammatical character. While the part-of-speech membership of a word does not give away the particu-

lar meaning, it does suggest the general type of that meaning, whether action, object, substance, or whatever. In learning a language, therefore, it must be useful to discover the semantic correlates of the various parts of speech; for this discovery enables the learner to use the part-of-speech membership of a new word as a first clue to its meaning. The present experiment with very young children who are learning English indicates that in this language, at least, the semantic implications of the verb, mass noun, and particular noun are discovered by native speakers. It now seems quite probable that speakers of other languages will also know about the semantics of their grammatical categories. Since these are strikingly different in unrelated languages, the speakers in question may have quite different cognitive categories. It remains to be determined how seriously and how generally thought is affected by these semantic distinctions.

#### SUMMARY

Descriptive linguistics defines the parts of speech in strictly formal or syntactical terms. Nevertheless, the parts of speech usually have distinct semantic characteristics. These characteristics do not hold for all members of the various parts of speech, however, and so cannot serve to define the parts of speech for the purposes of linguistic science. Human beings are generally adept at picking up imperfect probabilistic implications, and so it may be the case that native speakers detect the semantic nature of the parts of speech of their language. It was shown that the nouns used by young English-speaking children were more reliably the names of things and their verbs more reliably the names of actions than is the case for the nouns and verbs used by English-speaking adults. It was shown experimentally that young English-speaking children take the part-of-speech membership of a new word as a clue

to the meaning of the word. In this way, they make use of the semantic distinctiveness of the parts of speech. It seems likely that speakers of languages other than English will also have detected the semantic characters of their parts of speech. There is a sense, then, in which this grammatical feature of a language affects the cognition of those who speak the languages. Differences between languages in their parts of speech may be diagnostic of differences in the cognitive psychologies of those who use the languages.

#### REFERENCES

1. BROWN, R. W. Language and categories. Appendix to J. S. Bruner, J. J. Goodnow, & G. A. Austin, *A study of thinking*. New York: Wiley, 1956.
2. BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages. *J. abnorm. soc. Psychol.*, 1955, 50, 388-393.
3. BROWN, R. W., LEITER, R. A., & HILDUM, D. C. Metaphors from musical criticism. *J. abnorm. soc. Psychol.*, 54, 347-352.
4. BUCKINGHAM, B. R., & DOLCH, E. W. *A combined word list*. Boston: Ginn, 1936.
5. FRIES, C. C. *The structure of English*. New York: Harcourt Brace, 1952.
6. HOIJER, H. (Ed.) *Language in culture*. Chicago: Univer. of Chicago Press, 1954.
7. HORN, E. The commonest words in the spoken vocabulary of children up to and including six years of age. In National Society for the Study of Education, *Twenty-fourth yearbook*. Bloomington, Illinois: Public School Publishing Company, 1925.
8. INTERNATIONAL KINDERGARTEN UNION, CHILD STUDY COMMITTEE. *A study of the vocabulary of children before entering the first grade*. Washington, D. C.: The Int. Kindergarten Union, 1928.
9. LEE, DOROTHY D. Conceptual implications of an Indian language. *Phil. of Sci.*, 1938, 5, 89-102.
10. RINSLAND, H. D. *A basic vocabulary of elementary school children*. New York: Macmillan, 1945.
11. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bur. of Publ., Teachers College, Columbia Univer., 1944.
12. WHORF, B. L. *Language, thought, and reality: selected writings*. Cambridge: Technology Press, 1956.

Received August 17, 1956.



# A STUDY OF ANXIETY IN THE SOCIOPATHIC PERSONALITY<sup>1</sup>

DAVID T. LYKKEN

*University of Minnesota*

THE concept of the psychopathic personality includes so heterogeneous a group of behavior disorders as to be at least two steps removed from the level of useful psychiatric diagnosis. Sociopathic personality is a more recent designation (1) which refers to a subgroup of these disorders in which the pathognomic characteristics are impulsiveness, antisocial tendencies, immorality, and a seemingly self-destructive failure to modify this pattern of behavior in spite of repeated painful consequences. This category may be regarded as a genus composed of phenotypically similar, but etiologically distinct, subtypes such as the dissocial and the neurotic sociopaths.

A third species has been described (3, 12, 13, 14, 17), which may be called *primary sociopathy*, in which neither neurotic motivations, hereditary taint, nor dissocial nurture seem to be determining factors. Cleckley (3) has reported the chief clinical characteristic of this group as a lack of the normal affective accompaniments of experience. If this observation is correct, it would point the way toward accurate diagnostic isolation of primary sociopathy as well as guiding research into the question of its etiology. Classification according to the presence or absence of defective emotional reactivity, therefore, satisfies one criterion of useful diagnosis in that it shows promise of relationship to the as yet unknown origins of the disorders to be distinguished.

The other requirement for useful diagnosis is that the criteria of classification must be objective. Clinical assessment of the "normality of the affective accompaniments of experience" is subjective and unreliable. In consequence, Cleckley's work has had as yet little real impact on psychiatric practice. By expressing this putative defect of the primary sociopath in terms of the anxiety construct of experimental psychology (18, 19, 20, 21, 22),

it becomes susceptible to quantification and empirical test.

An experimental hypothesis may now be formulated. Among persons conventionally diagnosed as psychopathic personality, those who closely resemble the syndrome described by Cleckley are (a) clearly defective as compared to normals in their ability to develop (i.e. *condition*) anxiety, in the sense of an anticipatory emotional response to warning signals previously associated with nociceptive stimulation. Persons with such a defect would also be expected to show (b) abnormally little *manifest anxiety* in life situations normally conducive to this response, and to be (c) relatively incapable of *avoidance learning* under circumstances where such learning can only be effected through the mediation of the anxiety response.

## METHOD

### *The Sample*

The extreme heterogeneity, even on the crudest descriptive level, of persons diagnosed as psychopathic personalities in various clinical or institutional settings complicated the selection of an appropriate experimental sample. The institution psychologists<sup>2</sup> were given a list of 14 criteria drawn from Cleckley (3, pp. 355-392) and were asked to compare against these criteria those inmates diagnosed as psychopathic personality. Inmates who, in their opinion, best fitted the Cleckley prototype were listed as candidates for experimental Group I, the primary sociopathic group. Inmates who they felt did *not* meet the criteria in important respects were listed as candidates for experimental Group II, designated as the neurotic sociopathic group. In this selection process, the psychologists were asked to reaffirm the original diagnosis, discarding from consideration for either group those inmates who, in their present opinion, would not be diagnosed as psychopathic personality at all.

A control Group III of 15 "normals," roughly comparable in age, intelligence, and socioeconomic background, was selected from the University General College and a local high school.

Group I, composed of 12 males and 7 females, had a

<sup>1</sup> Drawn from a thesis submitted to the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is indebted to his adviser, Professor Ephraim Rosen, and to others whose assistance aided in the completion of this research.

<sup>2</sup> The writer is indebted to the administrators and to the psychologists of the Minnesota State Reformatory, St. Cloud, Minnesota; the Minnesota State Reformatory for Women, Shakopee, Minnesota; the State Home for Girls, Sauk Centre, Minnesota; and the St. Peter State Hospital, St. Peter, Minnesota.

mean age of 21.6 years ( $SD = 4.3$ ), and a mean IQ of 109.2 ( $SD = 10.7$ ). Group II included 13 males and 7 females, had a mean age of 24.5 years ( $SD = 5.4$ ), and a mean IQ of 104.5 ( $SD = 8.8$ ). For the 10 male and 5 female normals, the mean age was 19.07 ( $SD = 3.2$ ), and the mean IQ 100.4 ( $SD = 10.2$ ). None of these group differences were significant.

### *The Measures and Testing Procedure*

It was necessary to do the testing at the several institutions under varying conditions. In all cases, however, the apparatus was arranged on a large table, the experimenter on one side and the subject (*S*) seated comfortably opposite. The *S* was told that he was assisting in a psychological experiment having no bearing on his personal record and that his performance would be treated with strict anonymity. An attempt was made throughout to keep the testing on an informal basis.

As an indicant of manifest anxiety as referred to in hypothesis *b*, an "Anxiety Scale" was constructed expressly for this study to supplement the Taylor scale and Anxiety Index which appear to be more strongly loaded on a factor of neurotic self-description. In this new scale, each of the thirty-three items involves two statements of activities or occurrences, matched for general unpleasantness or undesirability according to a modified Thurstone scaling procedure utilizing 15 college student judges. One activity of each pair is unpleasant, presumably because of its frightening or embarrassing character (e.g., "making a parachute jump" or "knocking over a glass in a restaurant"). The paired activity is intended to be onerous but not frightening (e.g., "digging a big rubbish pit" or "cleaning up a spilled bottle of syrup"). The *S* is required to choose that member of each pair which he would prefer as a lesser of evils. The degree to which the "frightening" alternatives are rejected is interpreted as an index of the extent to which anxiety determines behavior choices within the range of life situations sampled by this test.

The booklet form of the MMPI was used and the answer sheets scored and *K*-corrected in the usual way (10). The Anxiety Index, or AI, was calculated according to the formula given by Welsh (23). The Heineman form (11) of the Taylor scale was given and scored by subtracting the number of "anxiety" items rejected as "least applies to me" from the number endorsed as "most applies to me."

An avoidance learning test was given to determine whether there were group differences in capacity to learn on the basis of anxiety reduction (hypothesis *c*). It involved an elaborate, electrically operated mental maze which the *S* was given 20 trials to learn (the "manifest task"). At each of the 20 choice points in this maze, choice of one of the 4 possible alternatives (always an error alternative) gave an electric shock. It was intended that social and ego rewards should reinforce performance in the manifest task. Performance on the "latent task," which was to avoid the shocked alternatives—to err instead on the unshocked alternatives—was presumably reinforced only through anxiety reduction.

The measure of anxiety conditionability (hypothesis *a*) employed the GSR as the dependent variable. A

shocking electrode was attached to *S*'s nondominant hand, the GSR electrodes being already in place on the dominant hand. The *S* was told that after the blindfold had been replaced, he was to sit as quietly as possible for the next 30 to 40 minutes, during which time he would periodically hear a buzzer (which was then demonstrated) and occasionally receive a brief electric shock. When the *S* was seated comfortably and relaxed insofar as possible, the recording apparatus was started and the conditioning series (CS) begun.

Two buzzers were used which were distinguishably different in timbre rather than in pitch, the difference being one not easily labeled (to minimize verbal mediation of a discrimination between them). Buzzer No. 1 was used as the CS and was the only one reinforced; buzzer No. 2 was used to test for generalization effects. In all cases, stimuli of the conditioning series were presented as soon as GSR activity from preceding stimuli had subsided, the intertrial interval being therefore not constant within or between *S*s, but averaging between 20 and 60 seconds. (This method of stimulus timing automatically eliminates temporal conditioning.) When turned on, the buzzers sounded for a period of 5 seconds, controlled by an automatic timer.

The reinforcing stimulus or UnCS was an electric shock from a 700-volt AC supply through two 68,000-ohm series resistors, presented automatically for about 100 milliseconds just before the termination of the CS (buzzer No. 1). The shock was applied between an electrode on the palm of one hand and the GSR ground electrode on the palmar tip of the middle finger of the opposite hand. The shock sensation was felt mainly on the richly innervated finger tip and was a decidedly unpleasant stimulus, producing in most cases a pronounced startle reaction and in all cases a strong GSR.

The sequence of trials or stimulus presentations was as follows:

1. To permit the adaptation of unconditioned GSR to the buzzers themselves, stimuli were first presented without shock reinforcement for a total of 10 trials in the order 2, 1, 2, 1, S, 2, 2, 1, 1, 1, 1. A single preliminary shock was given in the series at the point S, separated by at least 30 seconds from the buzzers occurring before and after it.

2. Seven consecutive shock-reinforced presentations of the CS were given as the conditioning series, followed by four more reinforcements interspersed with four unreinforced trials with buzzer No. 2 in the order 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 2, 1, 1.

3. A total of 24 extinction trials was then given, the two buzzer stimuli being presented in the order 1, 2, 1, 1, 2, 1, 2, 1, 1, 2, 1, 1, 1, 2, 2, 1, 2, 1, 1, 2, 1, 1, 1. Considering only the CS, buzzer No. 1, the series therefore consisted of 6 prereinforcement trials, 11 reinforced conditioning trials, and 16 extinction trials.

Skin resistance was measured by a modification of a circuit suggested by Flanders (6) which passed an electronically regulated constant DC current of 40 microamperes through *S*. The electrodes were curved discs of Monel metal, 15 mm. in diameter, applied to the palmar surface of the distal phalange of the first, second, and third fingers of the same hand. The skin surface was first scrubbed with alcohol and then coated with Sanborn electrode paste. The exosomatic current was applied between the first and third fingers, which were



also connected to the push-pull input grids of a Sanborn Model 126 DC amplifier, driving a Sanborn Model 127 recording milliammeter. The electrode on the second finger was connected to amplifier and external ground. The instrument was calibrated before each use and provided a linear record of resistance and resistance change, accurate to less than  $\pm 50$  ohms.

All GSRs were recorded in terms of resistance change. A variety of transformations was then applied and tested against the usual criteria of normality of distribution, correlation with basal resistance, and homogeneity of variance across people with respect to several test stimuli (2, 8, 9, 16). The result of this analysis was that each resistance change was expressed as the logarithm of the ratio of that change to the mean resistance change produced by the first six electric shocks. This unit expresses the galvanic CR as a proportion of the individual's UnCR and, for a conditioning study, seems quite appropriate for individual comparisons.

Three GSR indices were derived from the protocols of the conditioning series: (a) GSR Reactivity, which is the mean GSR to the CS during the fourth through seventh conditioning trials; (b) GSR Conditioning, which is equal to (a) minus the mean GSR to the last three preconditioning trials and the last three extinction trials (this index measures essentially the slope of the conditioning curve or the increment actually produced by the reinforced trials); (c) GSR Generalization, the ratio of the mean GSR to buzzer No. 2 during early extinction trials 18, 20, 21, 23 to the mean GSR to buzzer No. 1 during trials 17, 19, 22, 24.

The testing sequence was as follows: (a) Anxiety scale; (b) GSR Conditioning series; (c) Avoidance Learning test; (d) MMPI (given during the week following the foregoing individual testing); (e) Taylor Manifest Anxiety Scale, forced-choice form given later with the MMPI.

RESULTS AND DISCUSSION

Scores on all measures were converted for easier comparison to a standard score form with each distribution having a grand mean of

TABLE 1  
GROUP MEANS ON ALL MEASURES:  
SIGNIFICANCE TESTS\*

Measure	Group			d-Test prob.†
	I	II	III	
Taylor Scale	471	556	462	.01
Anxiety Index	472	557	464	.01
Anxiety Scale	470	511	529	.05
MMPI Pd-Scale	532	547	395	.05
Avoidance Learning	461	501	558	.01
GSR Reactivity	498	494	534	.05
GSR Conditioning	478	483	551	.05
Generalization	473	542	490	—

\* All measures converted to a scale having an over-all mean of 500 and SD of 100.

† Probabilities given are for significance of largest difference (e.g., III - I for GSR Conditioning). Significance test was Festsinger's distribution-free 'd' test (5).

500 and a standard deviation of 100. Group means on all measures, together with significance test results, are given in Table 1.

It would clearly be too much to expect of the judgments based upon the Cleckley criteria that they should have perfectly separated the psychopathic sample into a "primary" species in Group I, and a neurotic or dissocial species in Group II. That the separation was reasonably good, however, is supported by the finding that Group II scored significantly higher than the normals on the Taylor scale, a great deal of evidence having accumulated (4, 7, 15) to indicate that this scale is primarily a measure of neurotic maladjustment or neuroticism rather than of anxiety level or anxiety reactivity *per se*. On the MMPI Anxiety Index, which like the Taylor scale is unquestionably polydimensional with a heavy loading on neuroticism, Group II again has the highest mean, with Group I again only slightly higher than Group III.

In contrast, the Anxiety scale, which was designed for this study and which is not loaded on neuroticism and only negligibly correlated with the Taylor scale or the AI, separated the groups in a different order. On this test, the primary types of Group I show the least anxiety reactivity, significantly less than the normals, with Group II falling in between but rather nearer to the Group III mean. This result appears to support hypothesis *b* of this study, that the subset of primary sociopaths show abnormally little manifest anxiety, i.e., anxiety reactivity to the real-life anxiety stimuli referred to in the questionnaire.

Both sociopathic groups scored significantly higher than the normals on the *Pd* scale of the MMPI, but this measure, which differentiates at the phenotypic or genus level, does not distinguish between the types or species of sociopathy represented in Groups I and II.

Schedule difficulties unfortunately led to a reduction in the number of Ss to whom the avoidance learning test could be given. With nearly half of the total group, the available testing time was too short to cover all of the procedures; in such cases the avoidance test, requiring nearly an hour to give, was passed over. Even on the residual sample of 34 Ss, however, rather clear-cut differences exist. As a crude, overall index of avoidance learning, the avoidance scores (shock errors divided by

unshocked errors) were averaged for all but the first of the 20 trials; this is the basis of the mean scores entered under "avoidance" in Table 1. The distribution was reversed to make high values represent greater avoidance of the shock. It is impossible, of course, to summarize adequately a complex learning process by a single numerical index of this sort, but in spite of these limitations, it is striking that Group I (primaries) shows the least avoidance as expected, Group II (neurotics) next, and Group III (normals) the most. The Group I versus Group III, and Group II versus Group III differences are significant by Festinger's *d*-test (5), and the actual distribution of scores shows the groups to be remarkably well separated (only 17 per cent overlap between Groups I and III). This result supports hypothesis *c* of this study, that the primary sociopath demonstrates defective avoidance learning.

#### *Results of the GSR Conditioning Series*

Of all the tests employed here, principal emphasis should be laid on GSR conditioning. The various difficulties attending the interpretation of GSR data are well known, but one fact stands out with relative certainty: given certain necessary conditions, if an *S* does *not* produce a GSR to a stimulus, one can be sure that he has not "reacted emotionally" to that stimulus.

The two numerical indices which were derived as alternative ways of representing in a single value the conditioning indicated by the GSR protocols (anticipatory GSR to the buzzer after several pairings with shock) have already been described. As shown in Table 1, the group means are in the expected order on both indicants, with Group I significantly lower than Group III on GSR Reactivity and GSR Conditioning (.05 level, *d*-test).

A somewhat more meaningful comparison is obtained by contrasting the reactivity by trials for the three groups. Group I shows the least GSR reaction to the CS in 14 out of the 16 double trials. Group II is significantly higher (.02 level) than Group I at the end of the extinction trials. The positions of Group II and Group III interchange during the series with Group II beginning to show greater reactivity during the extinction trials, suggesting a perseveration (failure of extinction) of the anxiety response in the neurotic group.

This trend was tested for statistical reliability by correlating the differences between Group II and Group III with the ordinal position in the conditioning series at which the difference was taken. The quadrant sign test (24) shows this association to be significant at the .01 level. This result supports hypothesis *a* of this study, that the primary sociopath is defective in his ability to condition the anxiety response.

The generalization scores were leptokurtically distributed, the group differences being determined by a few deviant *Ss*. Group II shows the highest mean generalization score, but the differences are not significant.

#### SUMMARY

Forty-nine diagnosed psychopaths were divided into two groups according to the descriptive criteria of Cleckley. Fifteen normals served as controls. A battery of tests related to anxiety reactivity or anxiety conditionability were administered. As compared with normals, the Cleckley, or "primary" sociopaths, showed significantly less "anxiety" on a questionnaire device, less GSR reactivity to a "conditioned" stimulus associated with shock, and less avoidance of punished responses on a test of avoidance learning. The "neurotic" sociopaths scored significantly higher on the Taylor Anxiety Scale and on the Welsh Anxiety Index.

#### REFERENCES

1. AMER. PSYCHIATRIC ASSOC. *Diagnostic and statistical manual: mental disorders*. Washington, D.C.: American Psychiatric Assn., 1952.
2. BITTERMAN, M. E., & HOLTZMAN, W. H. Development of psychiatric screening of flying personnel. III. Conditioning and extinction of the GSR in relation to clinical evidence of anxiety. *USAF Sch. Aviat. Med.*, 1952, Proj. No. 21-37-002, Rep. No. 3, N. 232 p.
3. CLECKLEY, H. *The mask of sanity*. (2nd ed.) St. Louis: C. V. Mosby, 1950.
4. ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety Scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, 50, 135-137.
5. FESTINGER, L. The significance of the difference between means without reference to the frequency distribution function., *Psychometrika*, 1945, 11, 97-105.
6. FLANDERS, N. A. A circuit for the continuous measurement of palmar resistance. *Amer. J. Psychol.*, 1953, 66, 295-299.
7. FRANKS, C. Conditioning and personality: a study



- of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52**, 143-150.
8. HAGGARD, E. A. Experimental studies in affective processes. II. On the quantification and evaluation of "measured" changes in skin resistance. *J. exp. Psychol.*, 1945, **33**, 46-56.
  9. HAGGARD, E. A. On the application of analysis of variance to GSR data. I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, **39**, 378-392.
  10. HATHAWAY, S. R. *Supplementary manual for the MMPI. Part I, The K scale and its use*. New York: Psychological Corp., 1946.
  11. HEINEMAN, C. E. A forced choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447-454.
  12. KARPMAN, B. Psychopathic types: the symptomatic and the ideopathic. *J. crim. Psychopathol.*, 1941, **3**, 112-124.
  13. KARPMAN, B. The myth of the psychopathic personality. *Amer. J. Psychiat.*, 1948, **104**, 523-534.
  14. KARPMAN, B. Conscience in the psychopath: another version. *Amer. J. Orthopsychiat.*, 1948, **18**, 455-491.
  15. KERRICK, JEAN S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1955, **50**, 75-77.
  16. LACEY, O. L., & SIEGEL, P. S. An analysis of the unit of measurement of the galvanic skin responses. *J. exp. Psychol.*, 1949, **39**, 122-123.
  17. LIPPMAN, H. S. Psychopathic behavior in infants and children: a critical survey of existing concepts. *Amer. J. Orthopsychiat.*, 1951, **21**, 227-231.
  18. MAY, M. A. Experimentally acquired drives. *J. exp. Psychol.*, 1948, **38**, 66-77.
  19. MILLER, N. E. Studies of fear as an acquirable drive. I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89-101.
  20. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 435-472.
  21. MOWRER, O. H. A stimulus-response analysis of anxiety. *Psychol. Rev.*, 1939, **46**, 553-565.
  22. MOWRER, O. H. Anxiety reduction and learning. *J. exp. Psychol.*, 1940, **27**, 497-516.
  23. WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65-72.
  24. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.

Received July 5, 1956.

# A STUDY OF AFFECTIVE RESPONSIVENESS IN A LIE-DETECTION SITUATION<sup>1</sup>

JACK BLOCK

*Institute of Personality Assessment and Research, University of California, Berkeley*

SOME years ago, H. E. Jones (9) found a relatively low relationship between the intensity of the galvanic skin response (GSR) and the amount of overt emotional expression when both were measured simultaneously in nursery-school children. Although stimuli which on the average elicited strong autonomic responses were also found to evoke a great deal of observable emotional expression, for individual children the correlation over situations between GSR activation and outward emotionality was quite low.<sup>2</sup> Since the measures employed were highly reliable, the low relationship within individuals between internal and overt manifestations of emotions suggested that stable individual differences exist in the patterning of emotional expression.

This provocative finding was repeated and extended by Jones in a further study (10). Here, one hundred adolescents were available who, in the course of participating in a longitudinal study, had experienced eleven test situations where GSR recordings were gathered. The 20 individuals who were highest in the average magnitude of their GSRs and the 20 who were lowest were selected, and the independently rated overt behavioral characteristics of the two groups compared.

The "high reactives" were rated as less assertive, less animated, less talkative, and less attention-seeking than the "low reactives," all of these differences occurring at the .01 level

of significance. The high reactive group was also evaluated as more calm, more deliberative, more good-natured, more cooperative, and more responsible than the low reactives, who were judged as irritable, excitable, impulsive, and immature.

Jones suggested that the high reactives are introversive characters, while the low reactives represent the psychoanalytic concept of "impulse neurosis" (2). The process of socialization is the agent causing the child to shift from undifferentiated, externalizing modes of expression to complex, internalizing avenues of affect discharge.

It is important to note that Jones also observed the presence of two additional patterns of emotional expression. He identified a "generalizer" type, consisting of children who respond at both overt and internal levels simultaneously, and a "reciprocal" type, represented by children who sometimes respond overtly but with little GSR reactivity and who at other times evidence a highly reactive GSR but with little overt response.

The work by Jones has been summarized here because it is surprisingly little known and because it served as a springboard for the present experiment. His results have implications for conceptions of the nature and etiology of psychopathy and for a more general understanding of the place and process of affect in personality.

The present study aimed at an independent verification and extension of Jones's findings. It contrasts with the previous work in several ways: the subject sample used was older and probably more homogeneous with respect to intelligence, interests, and socioeconomic level; there was only one GSR recording session; the more stressful circumstance of lying was employed as the stimulus situation; and finally, objective test measures as well as ratings were available for comparison with GSR reactivity.

## METHOD

### *Subjects and Experimental Procedure*

Seventy male applicants to a medical school underwent an 18-hour assessment program at the Institute

<sup>1</sup> This study was supported in part by grants to the Institute of Personality Assessment and Research from the Rockefeller Foundation and the National Institute of Mental Health of the National Institutes of Health, U. S. Public Health Service. Assessment is a group endeavor, and the writer is pleased to record his debt to the staff of the Institute who participated in the program and who helped provide the data reported here. The writer is particularly grateful to Samuel Fillenbaum for his advice and assistance.

An abbreviated version of this paper was read at the meeting of the American Psychological Association at San Francisco, California, September, 1955.

<sup>2</sup> The correlation over the 22 situations between the group's average GSR and the average rating on overt-ness of response was .83 and .80 in the initial and retest experiments. For individual children, the correlation averaged .20 and .25 in the two separate tests.



of Personality Assessment and Research. In age they ranged from 19 to 32 years with a mean at 22. Intellectually, the group was at a very superior level. Their socioeconomic level was clearly middle class.

Each of the subjects (Ss) was observed in a wide variety of personality revealing situations, ranging from charades and stress interviews to standardized psychodramas and dinnertime conversation. One of the procedures each S experienced was the lie-detection situation. At his scheduled time, S entered the experimental room where he was greeted by the experimenter (E) and seated before the imposing apparatus. A pneumograph was attached around S's chest, his left thumb was inserted into a pulse pressure recording device, and electrodes were attached to his right hand. Only the GSR circuit of the apparatus was in fact operative during the experiment, the respiration and pulse transducers being employed for their dramatic effect only.<sup>3</sup>

The E conducted a short interview, inquiring about the origins of S's medical interests, his professional aspirations, the nature of his father's occupation, and so on, while an assistant adjusted the apparatus. S was then informed that the apparatus was a lie detector and that his skill as a liar was to be investigated.

The S selected a number ranging from 1 to 5 without informing E of his choice. The E then queried S a number of times over each possibility, employing a fixed but unsystematic order of interrogation. To each question, S was instructed to say "no," even where denial meant a lie. Questioning was varied in phrasing and accentuation so as to maintain the interpersonal aspects of the essentially standard stimulus situation. At the end of the series of 30 questions, S informed E as to his actual choice. In two additional replications, S selected an alternative (one of six months, one of six colors) and, again, E repeatedly interrogated him as to his choice. Again, S denied all possibilities, including the one he had in fact chosen. The temporal relations of stimuli and responses were recorded by noting E's questions on S's GSR record.

It was found that some Ss displayed a diffuse and unselective GSR responsivity and some were relatively unreactive; most were selectively responsive, the GSR tending to appear only after the lie. Over the three trials, an accommodation effect was usually observed as the S adapted to the situation.

Two raters judged the GSR records for amount of reactivity and selected the 20 most highly reactive individuals and the 20 least reactive individuals. Reactivity was judged primarily on the basis of the number of GSRs the subject manifested without further quantification of the size or duration of the response. Ratings were employed rather than a simple count because of the necessity to evaluate occasional movement and instrument artifacts. Classification of Ss into the categories was highly reliable, 36 of the 40 criterion cases being selected by both raters. The disagreements were resolved by mutual consent of the

raters. Categorization of Ss into the two GSR criterion groups occurred before the examination of any other data. The results to be reported in the next section are based upon the comparison of the two groups, Reactors and Nonreactors, on a number of measures.

## RESULTS

### *Intellectual Differences*

Intelligence is a powerful determiner of behavior, and its role in any particular context requires evaluation. Accordingly, the Reactors and Nonreactors were compared on measures defining various aspects of intellectual functioning. These included the Medical College Admissions test, grade-point average, a college vocabulary test (5), the Gottschaldt test (3), and a measure of originality based upon three of the Guilford tests of originality (6).

On none of these measures did the differences between the two groups approach significance. There is no evidence that such differences as may exist between the two groups can be ascribed to intellectual factors as these are ordinarily understood and measured.

### *Differences on Rating Measures*

At the end of assessment, each S was evaluated by each staff psychologist. One of the means employed for codifying staff impressions of each S was the Q-sort method of personality description (13); another was the Gough Adjective Check List (4).

A comprehensive 115-item Q set was used which included items oriented toward highly inferential features of the personality system as well as toward the more readily observable aspects of the individual.<sup>4</sup> Five psychologists, none of whom had observed the S's behavior in the lie-detection situation, Q sorted each S. These data, combined into a composite by equally weighting the ratings of each judge, codified the impressions and evaluations of these staff members with respect to the subject. Because the psychologist raters had no information as to the behavior of the S in the test situation, the Q sort data provide a totally independent source of information as to the personality characteristics of each S.

The comparison of the Q ratings for Reactors with the ratings for Nonreactors revealed twelve items differentiating at or beyond the

<sup>3</sup> Although there is some controversy among practitioners as to the suitability of the GSR as a lie detector under actual field conditions, there appears to be agreement as to its validity in experimental contexts (8, 11).

<sup>4</sup> A copy of the Q set employed is available from the writer.

.05 level of significance. These are listed in Table 1.

A separate group of seven staff psychologists employed the Adjective Check List in their characterization of each *S*. These data, when summed across raters and analyzed, served as a cross check on and amplification of the *Q* findings. The adjectives differentially characterizing the Reactors and Nonreactors at or beyond the .05 level of significance are presented in Table 2.

It is clear from the nature of the differentiating *Q* items and adjectives that the two groups differ along some version of the psychological dimension suggested by Jones. The Reactors are perceived as withdrawing, worrying individuals who turn their anxieties toward internal routes of expression. The Nonreactors are evaluated as independent, aggressively direct, and relatively nonconforming, all of these being visible rather than inward or covert expressions of impulses.

#### *Inventory Items Differentiating the Groups*

A number of personality inventory items (from the Minnesota Multiphasic Personality Inventory and the California Psychological Inventory) distinguished between the two groups. Summarizing their import, it would appear that the Reactors introspect relatively often, are sometimes radically subject to their moods and impulses, and are very much concerned with the appropriateness of their behavior in social contexts. In contrast, the Nonreactors indicate an absence of rumination and a self-control of their behavior. Also expressed is a preference for autonomy and a difficulty with representatives of authority.

#### *Some Test Measures Differentiating the Groups*

All *Ss* experienced the rod-and-frame procedure developed by Witkin (14). Here the *S* is placed in a totally dark room before a luminous square tilted from the vertical. He revolves, via a suitable control, a rod centered within the frame until he perceives the rod as at the true vertical (or horizontal). In the blackness of the experimental room, the only external cue to verticality is the misleading frame. An alternative basis for judgment is available from internal, proprioceptive gravitational cues. In the present study, the Reactors proved to be more affected by the ex-

TABLE 1  
Q ITEMS DISTINGUISHING BETWEEN THE REACTOR  
AND NONREACTOR GROUPS

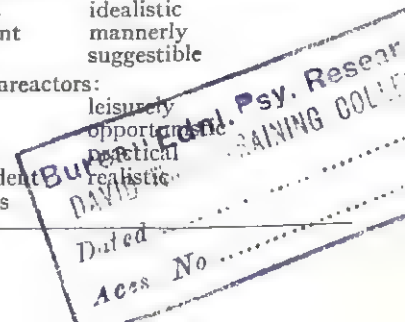
Characteristic of Reactors:	
10. His anxiety and tension find outlet in bodily symptoms and dysfunction; anxiety and tension are converted into somatic symptoms	.05
11. Is protective of those close to him	.05
14. Is readily dominated by others; submissive	.05
19. Seeks reassurance from others	.05
30. Gives up and withdraws where possible in the face of frustration and adversity	.05
40. Is vulnerable to real or fancied threat; generally fearful; is a worrier	.05
70. Behaves in an ethically consistent manner	.05
74. Lets other people take advantage of him; allows exploitation	.05
Characteristic of Nonreactors:	
94. Expresses his hostilities directly	.01
96. Values his own independence and autonomy	.01
1. Critical, not easily impressed, skeptical	.05
62. Tends to be rebellious and non-conforming	.05

ternal frame of reference than were the Nonreactors ( $p < .05$ ). That is, the Reactors were more externally oriented and more open to outside influence in the presence of uncertainty than were the Nonreactors, who lent greater credence to their proprioceptive cues.

A rather curious result emerges from a comparison of the behavior of the two groups in the Crutchfield independence-of-judgment experimental procedure (1). Here the subject is exposed to a group consensus about a problem which differs from what the *S* perceives as the correct or preferred solution. The extent to which *S* conforms to the group opinion in his manifested choice provides a measure of independence found by Crutchfield to be richly related to character structure.

TABLE 2  
ADJECTIVES DISTINGUISHING BETWEEN THE REACTOR  
AND NONREACTOR GROUPS  
(.05 level of significance)

Characteristic of Reactors:	
cautious	idealistic
dependent	mannerly
dreamy	suggestible
Characteristic of Nonreactors:	
clever	leisurely
cool	opportunistic
evasive	practical
independent	realistic
ingenious	





In the present analysis, Reactors and Nonreactors did not differ in amount of independence as expressed by their choice behavior. But subjects also expressed their subjective confidence in their judgments. The Reactors, although actually conforming no more than Nonreactors, proved to be less confident about their course of action ( $p < .001$ ). Independence (or conformity) in the Reactor thus appears to have a different psychological base than in the Nonreactor. It might even be said that manifesting independence in the face of uncertainty and doubts, as did the Reactors, required more courage than equivalent behavior where subjective doubt does not prevail, as among the Nonreactors.

Although the classification into Reactors and Nonreactors was based solely upon the evaluation of the GSR, some evidence for the physiological generality of reactivity may be adduced from a subsequent finding. Following a stressful psychodramatic situation (7), it was found that the systolic blood pressure of Reactors was raised significantly more than that of the Nonreactors ( $p < .01$ ).

#### DISCUSSION

The present results, taken together with those of Jones and others, provide reasonable evidence that electrodermal responsiveness is related to personality structure and behavior. While in both studies, the number of significant relationships is small considering the multiple tests of significance, the consonance of both sets of data argues for their psychological stability. The placing of these data within a conceptual frame of reference is another matter.

An interpretation of the findings in terms of the notion of externalization-internalization simply does not serve. The externalizing-internalizing dimension has been only informally conceptualized, but it appears to be concerned with whether need tensions are discharged via external, action modes of expression or whether needs are routed internally into cognitive and visceral channels of discharge. Implicit here is the assumption that there exists a reciprocal relationship between overt motor response and cerebral or autonomic activity. That is, overt motor response *displaces* internal activity and *vice versa*. Jones's finding, supported by Seymour (12), that there are "generalizers," individuals whose expression of need states is

simultaneously overt and covert, upsets such a notion. Observations of a culture like the Italian, where emotion is vigorously displayed but also apparently internally experienced, also weigh against the reciprocal theory of externalization-internalization. Clearly, an additional dimension or a reformulation of the concept is required in order to encompass the various patterns of emotional expression which have been noted.

It is suggested here that an important distinction is being obscured in the definition of externalization-internalization by opposing overt motor response on the one hand against cognitive *and* visceral channels of discharge on the other. The italicization of *and* is important, for it is precisely this conjunction which may be inappropriate. From an instrumental point of view, cognitive activity *and* motor behavior are both means to drive reduction, whereas autonomic, visceral reaction serves no such function. Autonomic activity appears instead to be an expression of an affective state and does not in and of itself bring about a significant reduction of the disturbing emotional charge. Tension reduction is brought about only by appropriate cognition, fantasy, or motor behavior, not by autonomic arousal.

If this reformulation appears a tenable one, then externalizing-internalizing can be redefined as a way of characterizing the direction or orientation of an individual's response to the impelling affective disequilibrium. GSR lability, from this point of view, is simply a convenient index of the tensional state, of whether an individual has anxiety or affect in a given situation or whether he does not have anxiety or emotion in that situation. It has no implication, however, for the direction of the response precipitated by the emotion. Since affective responsiveness may eventuate in motor as well as cognitive behavior, the dilemma previously posed by the "generalizer" no longer exists.

The dimension distinguishing between the Reactors and Nonreactors in the present study, may be the dimension of affective responsiveness without regard for just how the individual copes with emotional states. There has been little direct investigation of affective responsiveness although it is known that there are great differences in the extent to which individuals experience affect and have their self-

percept brought forward into consciousness as a conditioner of behavior. Some individuals experience too much affect, e.g., anxiety neurotics, and are thereby incapacitated. But others do not experience sufficient emotion, sufficient in relation to a normative criterion, and therefore do not react on the basis of an implicit but culturally shared set of behavioral premises. Because they operate from a different premise system, these individuals are more likely to emit behavior which is inconceivable to the affectively responsive observer. Such people include the primary psychopath, certain forms or stages of the schizophrenic reaction and, in general, those individuals identified by psychiatrists as possessing "flattened affect."

## SUMMARY

Seventy Ss, during the course of an assessment program, experienced a lie-detection situation where GSR reactions were recorded. The twenty most reactive subjects and the twenty least reactive subjects were selected and compared on a variety of measures. Reactors appeared to be more dependent, dreamy, idealistic, and suggestible; Nonreactors were evaluated as relatively cool, evasive, opportunistic, and independent. The findings were related to the previous work of Jones, and a reformulation of the notion of externalization-internalization was offered.

## REFERENCES

1. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191-198.
2. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
3. GOTTSCHALDT, K. Über den einfluss der erfahrung auf die wahrnehmung von figuren. *Psych. Forsch.*, 1926, **8**, 261-317; 1929, **12**, 1-87.
4. GOUGH, H. G. *The adjective check list*. Berkeley: Univer. California Printing Department, 1952.
5. GOUGH, H. G., & SAMPSON, H. *The college vocabulary test*. (Form A). Berkeley: Univer. of California Inst. of Personality Assessment and Research, 1954.
6. GUILFORD, J. P., WILSON, R. C., CHRISTENSEN, P. R., & LEWIS, D. J. A factor-analytic study of creative thinking: I. Hypotheses and descriptions of tests. *Reports from the Psychol. Lab.*, No. 3, Los Angeles: Univer. of Southern California, 1951.
7. HARRIS, R. E., SOKOLOW, M., CARPENTER, L. G., FREEDMAN, M., & HUNT, S. P. Response to psychologic stress in persons who are potentially hypertensive. *Circulation*, 1953, **7**, 874-879.
8. INBAU, F. E. *Lie detection and criminal interrogation*. (2nd ed.) Baltimore: Williams & Wilkins, 1948.
9. JONES, H. E. The galvanic skin response as related to overt emotional expression. *Amer. J. Psychol.*, 1935, **47**, 241-251.
10. JONES, H. E. The study of patterns of emotional expression. In M. L. Reymert (Ed.), *Feelings and emotions: the Mooseheart Symposium*. New York: McGraw-Hill, 1950.
11. MACNITT, R. D. In defense of the electrodermal response and cardiac amplitude as measures of deception. *J. crim. Law and Criminol.*, 1942, **33**, 266-275.
12. SEYMOUR, R. B. Personality correlates of electrodermal resistance and response. Unpublished doctor's dissertation, Univer. of California, 1950.
13. STEPHENSON, W. *The study of behavior*. Chicago: Univer. of Chicago Press, 1953.
14. WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

Received July 24, 1956.



# THE IMMEDIATE AND DELAYED EFFECT OF FAILURE AS A FUNCTION OF TASK COMPLEXITY AND PERSONALIZATION OF FAILURE

CHARLES B. TRUAX AND BARCLAY MARTIN

*University of Wisconsin*

PREVIOUS research has suggested various factors as being important in determining the effects of experimentally induced stress upon performance. This study is concerned with the manipulation of several of these factors within the framework of one experimental design. There is ample evidence that individual differences in general anxiety level, whether measured by paper-and-pencil tests or assessed by psychiatric diagnosis, are related to performance under stress conditions (1, 3, 4, 5, 9, 14, 15). The typical finding in this respect is an interaction of anxiety level by stress, so that high anxious subjects (*Ss*) tend to do less well under stress than under neutral conditions, and low anxious *Ss* tend to perform relatively better under stress than under neutral conditions.

This finding, however, cannot be accepted without qualification, and it was essentially the purpose of this study to investigate some of the factors which may affect this interaction of anxiety level and stress. One such factor is the level of complexity of the task being performed or learned. Studies have generally indicated that the more complex the task, the more poorly the high anxiety *Ss* do in relation to the low anxiety *Ss* (2, 7, 10, 12, 13). Thus, in the present experiment two levels of complexity have been introduced.

Another factor which may influence the effects of failure (as one form of stress) on performance is the length of the time interval interposed between the failure and the post-failure performance. This variable has usually been described in terms of the immediate versus delayed effects of failure, and previous findings are somewhat conflicting. Russell (8) found failure conditions to be associated with poorer retention scores than nonfailure conditions under immediate, but not delayed (24 hours), recall conditions. The task used was a list of 12 nonsense syllables learned by the serial anticipation method. Farber, Russell, and Andreas (1), using a paired-associate verbal learning task, found that failure adversely affected performance on immediate tests of

retention; but on delayed tests, the failed group tended to perform better than the controls. Only the decrement under the immediate testing conditions was significantly different between the failed and nonfailed groups however.

On the other hand, Zeller (16, 17) found both an immediate and delayed (72 hours) effect of failure relative to nonfailure conditions, using both a recall and a relearning score based on a paired-associate task. One clue suggesting an explanation for the discrepancy between Zeller's study and the others mentioned above may be found in the nature of the failure comments given. Zeller's failure remarks were without a doubt extremely severe and were to a considerable extent personally insulting to *S*. The comments in the Russell and the Farber, Russell, and Andreas studies, while without question indicating to *S* that he failed, were clearly much less severe and personalized. It may be, then, that delayed effects of failure become apparent only under rather strongly personalized conditions. In other words, failure experiences which are explicitly made to reflect upon the general competence and integrity of *S* may have longer lasting effects than those which imply only failure on the immediate performance.

To explore these possibilities, immediate and delayed postfailure testing, mild and severe personalization of failure, two degrees of task complexity, and subgroups divided at their medians on the basis of either the Taylor Manifest Anxiety Scale or the Stroop Interference Test were all incorporated as variables within the experimental design. The Stroop test (11) was included because it seemed likely that it might give some measure of a person's general susceptibility to interfering response tendencies, a trait that Waterhouse and Child (14) have suggested as being closely related to quality of performance under stress.

## METHOD

The *Ss* were 128 female students in the introductory psychology course at the University of Wisconsin. Sixteen *Ss* were randomly assigned to each of 8 treatment

groups. Since the two different tasks involved different dependent variables, the results had to be analyzed separately; accordingly, the study can be described as consisting of two separate two-by-two factorial designs. Thus, for the simple task condition, there were two levels of personalization (mild and severe) of failure orthogonal to two intervals between pre- and postfailure testing (20 seconds and 24 hours). In the complex task condition, a similar design was used. The individual difference measure was introduced by dividing the subgroups at the median on either the Taylor Manifest Anxiety Scale or the Stroop Interference Test, adding an additional dimension to the factorial design. Since Taylor scores were not obtained on several Ss, it was necessary to reduce all subgroups to an  $N$  of 14 in order to retain cell proportionality for this measure.  $N$ s of 16 were used for the Stroop test. The Taylor Manifest Anxiety Scale was administered to the Ss as a group in their regular class period before the individual testing session.

The Stroop Interference Test was given at the beginning of the individual session. This test consists of three parts: Card A, a page on which only the four color names of red, blue, yellow, and green are printed in black ink in a random arrangement of eleven rows of ten words each; Card B, a page on which the same colors appear in  $\frac{3}{8}$  inch colored dots, arranged randomly in the same format as Card A; and Card C, a page on which the same colors are printed in random order, with each color name printed randomly in a color other than the one it designates. The S's task was (a) first to read as fast as possible the color names on Card A, (b) to name as fast as possible the colors of the color dots on Card B, and (c) to name as fast as possible the colors on Card C in which the words were inked, ignoring any interference produced by the "wrong" names. The score used was the ratio of the time to finish Card B to the time to finish Card C.

*The simple task.* The simple task consisted of S's adding 78 pairs of one-digit numbers. The score was the time in seconds to complete the problems. Uncorrected errors on this extremely simple task were not frequent enough to affect the time score. The same task was used in the postfailure testing.

*The complex task.* The complex task consisted of the Arithmetic subtests of the Wechsler-Bellevue Intelligence Scale, Forms I and II. One form was presented under prefailure, and the second form was given under postfailure conditions with order counterbalanced within each of the 8 major experimental groups. The weighted score derived from the Wechsler-Bellevue standardization norms, involving both time and accuracy measures, was used.

*The failure inducing task.* Failure comments were given for a totally different task so that all Ss would experience failure in a common situation. This task was an extended version of the Wechsler-Bellevue Digit Symbol Substitution Test. It was presented on a dittoed sheet with the label, "Wisconsin Digit Symbol Test of Intellectual Ability, College Form A," prominently displayed across the upper third of the page. The time allotted for the completion of this task was varied with S's speed to insure that S would complete only three-fourths of the test.

*Procedure and instructions.* The following instructions were given initially for the Stroop Interference Test:

We are in the process of obtaining normative data on a test which we hope will be a good measure of Reading Ability. This test is in three parts, and we'll start off with the first part now.

The Stroop test was then administered, after which S was told,

As you may know, in standardizing a test you have to get other measures on the subjects in order to check for a normal distribution. For example, if all of our subjects were exceptionally intelligent or quick reacting, the Reading Test wouldn't be very good for estimating the ability of the average college student. So we have some other tests for you to take.

The simple or complex task was then administered, depending on the treatment group to which S was assigned. After this task, the following "explanation" was given of the failure inducing task:

Reading ability, as you probably know, is related to IQ, so to standardize the Reading Test for a college population, we will have to get a measure of the IQ of the subjects used in this study. We will use the Wisconsin Digit Symbol Test. It's one of the shortest IQ tests available, but it's still very reliable. The average college student will get about 114; the brightest student so far in this study has got 129 items right.

Don't expect to finish the test entirely in the two minutes you will have—just do the best you can.

After the failure inducing task was administered, experimenter (E) laid a template over the test and pretended to score it. The following comments (which had previously been memorized by E) were then made in the "severe personalization" group:

Damn! (E frowns and shakes head). Ninety-two. You really goofed this test . . . . (S's first name). You're well below the norm of 114. (Pause.) You having much trouble keeping up with your classes? Well, either you were just goofing up this experiment . . . . (S's first name), or something's wrong with you. A score of 92! Damn!!!

The following comments (which had previously been memorized by E) were made in the "mild personalization" group:

Ninety-two. (E pauses and continues in a moderate tone.) The score is 92. (Pause.) This is below the norm and not very good. (Pause.) A score of 92!

E paced the mild and severe personalization remarks so that the same amount of time was taken in giving both, about 17–18 seconds.

After the induction of failure, half the Ss were retested immediately on the performance task (either simple or complex), and an appointment was made for the other half to meet for a second session 24 hours later. All Ss were given the following instructions before the administration of the postfailure test:

Well, let's do this arithmetic (addition) test over again. Remember, I'll be timing you, so work as quickly as possible without making mistakes.

At the conclusion of the experiment, each S was told the purpose of the study and allowed to vent her feelings about the failure experience.

## RESULTS

*Simple task.* There was a negative correlation of  $-.68$  between scores on the simple task



under the prefailure condition and the change score from pre- to postfailure. Also, by chance apparently, the mild and severe personalization groups differed greatly in their prefailure scores. Accordingly, analysis of covariance was used, making it possible to analyze only that variance in the change scores that was uncorrelated with the prefailure scores. The variance within the different treatment groups did not differ significantly as indicated by Bartlett's test for homogeneity of variance.

It should be mentioned, before these analyses are considered, that a previously performed analysis of variance with repeated measures indicated a significant tendency ( $p < .001$ ) for all Ss to improve their performance on this simple task after failure. This general tendency would not appear, of course, in the analysis of covariance in which change scores were used; but it indicates that any found effects will represent differential tendencies to improve from pre- to postfailure conditions.

The Stroop test did not prove to be significantly related to change in performance on either the simple or complex tasks, and results with respect to this measure were not further considered.

Table 1 presents the analysis of covariance with the subgroups divided into high and low Taylor scale scores. The main effect of the immediate-delay variable was significant in the direction of greater improvement in performance under the delayed postfailure testing. Individual differences in anxiety were also significantly related to change in performance. High anxious Ss showed greater improvement ( $p < .01$ ) than low anxious Ss, and there was an interaction of anxiety with mild-severe

TABLE 1  
ANALYSIS OF COVARIANCE FOR THE SIMPLE  
TASK USING THE TAYLOR SCALE

Source	df	Mean Square	F
Taylor scale (A)	1	110.94	7.25**
Personalization (B)	1	1.87	—
Immediate-Delay (C)	1	76.80	5.02*
A × B	1	203.44	13.24***
A × C	1	41.89	2.74
B × C	1	15.47	1.01
A × B × C	1	7.22	—
Within groups	47	15.31	—

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

TABLE 2  
ANALYSIS OF COVARIANCE FOR THE COMPLEX  
TASK USING THE TAYLOR SCALE

Source	df	Mean Square	F
Taylor scale (A)	1	9.05	2.86
Personalization (B)	1	0.01	—
Immediate-Delay (C)	1	30.93	9.76**
A × B	1	0.49	—
A × C	1	0.05	—
B × C	1	1.87	—
A × B × C	1	12.09	3.82
Within groups	47	3.17	—

\*\*  $p < .01$ .

personalization ( $p < .001$ ), suggesting that high anxious Ss improved *less* under severe personalization than under mild, whereas low anxious Ss improved *more* under severe than under mild personalization.

This interaction was further investigated by performing an analysis of covariance upon the high and low groups separately. The low anxiety group was found to perform significantly better under severe than mild personalization ( $p < .05$ ), but there was no evidence that the high anxiety group performed better under mild than severe personalization. The interaction effect was apparently confined to the lower portion of the anxiety distribution.

*Complex task.* A repeated measures analysis of variance revealed no significant tendency for Ss in general to change in performance from pre- to postfailure on the complex task. Initial scores correlated with change scores  $-.54$ , so analysis of covariance was also used here. Table 2 presents the analysis of covariance for subgroups divided on the basis of the Taylor scale. The immediate-delay variable was significant ( $p < .01$ ), with Ss again showing more improvement in performance under the delayed postfailure testing. No other variables were significant.

It is sometimes reported that variability increases as a function of failure (6). The variability in the prefailure situation was compared to that in the postfailure situation for both simple and complex tasks. Changes were small and insignificant in both cases; the  $F$  values were, respectively, 1.19 and 1.02.

The differences in results between the simple and complex tasks do not appear to be due to differences in the extremeness of groups based on the Taylor scale. Table 3 presents the mean Taylor scale scores for the groups and their

TABLE 3  
MEANS AND PERCENTILE RANKS OF GROUPS  
ON TAYLOR SCALE

Group	Simple Task		Complex Task	
	Mean	P.R.	Mean	P.R.
High anxiety	23.22	86	20.95	80
Low anxiety	9.72	32	10.00	34

percentile rank as estimated from the University of Iowa normative data.

### DISCUSSION

No main or interactive effects were found when the Stroop Interference Test was used as the individual difference measure. These negative results suggest that anxiety is either not fruitfully conceptualized in terms of a general susceptibility to interfering responses as Waterhouse and Child (14) suggest or that the Stroop test is not a valid measure of this trait.

When the Taylor scale is used to assess individual differences, significant relationships were found with pre- to postfailure performance changes for the simple but not for the complex task. This suggests that the Wechsler Arithmetic Subscale may be relatively less sensitive to the effects of failure than the simple additions task, although there was a tendency for performance on the Arithmetic subscale to be affected by the immediate-delay variable. Both tasks involve relatively well learned past habits, and this common aspect of both tasks is probably not particularly susceptible to the effects of failure and general anxiety level. The simple additions test, however, is much more predominantly a speed test, and it may be that its relationship to the failure and anxiety variables is essentially a function of the speed factor, which seems much less important in the Wechsler problems.

The significant effects of the immediate-delay variable for both the simple task and the complex task suggest that failure has a greater effect after a 24-hour delay than under immediate retesting conditions. Nevertheless, this interpretation requires considerable caution since, first, it may hold only for conditions in which performance improves after failure, and second, no control for practice effects was available in this experiment. With regard to the latter consideration, however, the contention would have to be that practice has a

more beneficial effect after a 24-hour delay than upon immediate retesting, which seems possible but rather improbable.

The results of this study are in some respects consistent with those of Farber, Russell, and Andreas (1), who found that a failed group did worse than a nonfailed control group under immediate retesting conditions, but tended to do better than the control group under delayed retesting. Russell (8) also found the negative effects of failure to be greater in the immediate than in the delayed situation. With the present results, these findings suggest that the detrimental effects of failure are more pronounced in the immediate than in the delayed situation, whereas the facilitating effects of failure are more pronounced in the delayed than the immediate situation.

For the simple task, it was found that high anxious Ss showed greater improvement after failure than low anxious Ss. This result appears to contradict the usual findings. Williams (16), employing a similar task of simple addition problems, but using the Waterhouse and Child Interfering Tendency Questionnaire to select high and low groups, found an interaction in the opposite direction ( $p < .10$ ), with high scorers performing about the same after non-failure as after failure conditions. Waterhouse and Child (14) report similar interactions with somewhat different kinds of tasks.

However, if one introduces the variable of mild-severe personalization of failure, the results of the present study no longer seem so contradictory of previous findings. The low anxious Ss show more improvement under severe than mild personalization of failure, whereas the high anxious Ss do not show this differential improvement. The resulting interaction of anxiety with mild-severe personalization therefore seems to represent a psychologically meaningful process. Thus, it looks as though the performance of high anxiety Ss on this relatively simple task is not adversely affected by either mild or severe failure. On the contrary, their performance is comparably facilitated by the additional anxiety or motivation under both conditions. The low anxiety Ss, however, show a greater improvement under severe than mild personalization. It may be that high anxiety Ss are approaching an optimum state of tension in the mild conditions and, accordingly, further stress does not enhance performance; but the low anxious Ss,



having not yet reached the optimum level of tension, continue to improve in performance under severe personalization. This finding is consistent with the general hypothesis that adequacy of performance as a function of anxiety is a curvilinear function of an inverted U-shape type. The fact that the high anxiety group has apparently not gone beyond the optimum state of tension to the point where performance begins to be disrupted may be due to the extremely simple nature of the one-digit additions task. This interaction of anxiety level with degree of personalization also indicates the importance of reporting in detail the operations defining the failure in any given piece of research, since this variable seems to affect performance significantly.

#### SUMMARY

This experiment was designed to study the effects of several variables upon changes in performance from pre- to postfailure conditions. All Ss were female college students. For one group of Ss performance was measured on a simple task, one-digit addition problems; for another group of Ss, performance was measured on a more complex task, the Arithmetic subscale of the Wechsler-Bellevue. A factorial design was used for each of these groups with the independent variables being mild-severe personalization of failure, immediate-delay of postfailure retesting, and high and low groups based on either the Stroop Interference Test or the Taylor scale.

All Ss tended to improve in performance after failure on the simple task. High anxious Ss tended to improve more than low anxious Ss. The interaction of anxiety with mild-severe personalization was significant with low anxious Ss tending to improve more under severe than mild personalization, and high anxious Ss improving about equally under both conditions. The immediate-delay variable was significant with Ss showing greater improvement after the 24-hour delay.

The only relationship found, when performance on the complex task was used as the dependent variable, was a tendency for the Ss to improve more after the 24-hour delay than upon immediate retesting.

The findings in regard to the immediate delay variable suggest the tentative hypothesis that the facilitative effects of failure are greater

after a delay, but the detrimental effects are greater when retesting is done immediately.

#### REFERENCES

1. FARBER, I. E., RUSSELL, W. A., & ANDREAS, B. G. The effect of failure upon performance in verbal and motor learning situations: A review of work done at the State Univ. of Iowa. Paper read at Midwest. Psychol. Ass., Chicago, April, 1949.
2. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120-125.
3. KAYE, D., KIRSCHNER, P., & MANDLER, G. The effects of test anxiety on memory span in a group test situation. *J. consult. Psychol.*, 1953, **17**, 265-266.
4. LUCAS, J. D. The interactive effects of anxiety, failure, and intra-serial duplication, *Amer. J. Psychol.*, 1952, **65**, 59-66.
5. MALMO, R. S., & AMSEL, A. Anxiety-produced interference in serial rote learning with observations on rote learning after partial frontal lobectomy. *J. exp. Psychol.*, 1948, **38**, 89-101.
6. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166-173.
7. MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91-96.
8. RUSSELL, W. A. Retention of verbal material as a function of motivation instructions and experimentally induced failure. *J. exp. Psychol.*, 1952, **43**, 207-216.
9. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561-565.
10. SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *The Kentucky symposium, Learning theory, personality theory, and clinical research*. New York: Wiley, 1953.
11. STROOP, J. R. Studies of interference in serial verbal reactions. *J. exp. Psychol.*, 1935, **18**, 643-661.
12. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.
13. TAYLOR, JANET A., & CHAPMAN, J. P. Paired associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68**, 671.
14. WATERHOUSE, I. K., & CHILD, I. L. Frustration and the quality of performance: III. An experimental study. *J. Pers.*, 1954, **21**, 298-311.
15. WILLIAMS, J. E. Mode of failure, interference tendencies, and achievement imagery. *J. abnorm. soc. Psychol.*, 1955, **51**, 573-580.
16. ZELLER, A. F. An experimental analogue of repression: II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, **40**, 411-422.
17. ZELLER, A. F. An experimental analogue of repression: III. The effect of induced failure and success on memory measured by recall. *J. exp. Psychol.*, 1951, **42**, 32-38.

Received July 30, 1956.

8-424  
8-370

## USES OF COMMON OBJECTS AS INDICATORS OF CULTURAL ORIENTATIONS<sup>1</sup>

WAYNE DENNIS

*Brooklyn College*

IF children are asked with regard to each of several objects, "What is it for?" one finds that children's concepts of the functions of identical things vary from society to society. Those functions of common objects which are frequently mentioned by the children of a social group, it is here proposed, are indicative of the activities, interests, and values that prevail in that society. That is, it is held that children value things for the uses to which they may be put, and, conversely, that frequently-mentioned uses are indicative of values. For example, if children in one society often mention boys and girls, cats and dogs, trees and sand, as functioning chiefly in play whereas the children of another group mention them more often in connection with work, these facts suggest that there is a difference between the two groups in regard to work-play emphases. The present paper illustrates the application of this method, The Uses Test, with three groups of children, an American group, a Lebanese group, and a Sudanese group.

If the objects used in questioning children are universally present, this test procedure can be employed in all cultures. Techniques for the study of child behavior that are universally applicable are very small in number. The method employed in the present study is offered as a useful addition to the small stock of tools available for cross-cultural research. It is believed to be particularly useful as an exploratory technique in societies whose patterns of

child behavior are but slightly known to the investigator.

### METHOD

The question employed is the simple and probably universal one, "What is ——— for?". Naturally, this question may be asked in regard to any plant or animal, any class of persons, or any part of the body.

The question, "What is ——— for?", was chosen because numerous studies indicate that the young child readily thinks in terms of use. He spontaneously asks what things are for, he defines objects in terms of their use, and he seems to presume that every object in the world has a use. No difficulties have been encountered in obtaining answers to the question from hundreds of children between ages five and ten in a variety of cultures. In all likelihood it can be used above and below the age levels represented in this study. The technique is not limited to the particular word list employed here. The objects concerning which children are questioned will vary with the interests and hypotheses of the investigator.

In the present study, the following words were used in the order given: the mouth, the hands, a mother, a father, a boy, a girl, trees, wood, dogs, cats, birds, stones, the sun, sand, rain, and gold. In the case of the American and Lebanese subjects, data were gathered by individual interview. The Sudanese subjects were instructed orally as a group, and they were requested to write their answers.

Although the question put to the child did not ask for more than one use for each object, some children gave more than one. In the present study, responses beyond the first were not included in the analysis of the data, unless specific reference is made to answers beyond the first one.

### Subjects

*American.* These were American children attending the American Community School in Beirut, Lebanon. The majority of these children were born in the United States and spent their pre-school years there. They were in Lebanon in 1955-56 because their parents were temporary residents of that country. Most of the fathers were employed by the University, or American business firms. Nearly all could be classified as belonging to the American middle class. All children between ages 5.0 and 11.0 were tested. These totaled 120 children, approximately equally representing the two sexes and the various ages.

*Lebanese.* This group consisted of an analytic sample of native-born children in Beirut, Lebanon. The sample was divided into 36 cells, produced by cross classifying in terms of four criteria: sex, private schools vs. public schools, age (5-6 years, 7-8 years, 9-10 years) and religious community (Arab-Christians, Moslems, and

<sup>1</sup> The research here reported was done during 1955-56 when the author was a visiting professor at the American University of Beirut, Lebanon. Expenses of the study were defrayed by a grant of the Rockefeller Brothers Fund to the American University of Beirut. The writer wishes to express his gratitude to Mrs. Adele Hamdan Taky Din, Miss Leila Biksmati, Mrs. Marie Therese Broussalian, Mrs. Yvonne Sayegh and Miss Aimee Sabbagh who served as research assistants, and to the principals, teachers, and pupils whose cooperation made the study possible. He is also greatly indebted to staff members of the departments of psychology and education, and to members of the child development seminar at the American University of Beirut, who gave invaluable assistance and counsel throughout the course of the investigation.



Armenians). With 25 children in each cell, the sample comprised 900 subjects in all. All children of appropriate ages were interviewed in each of several schools, the pupils being subsequently assigned to cells. If more than 25 so obtained Ss fell into a given cell, the number was reduced to 25 by discarding the data from Ss selected by chance.

*Sudanese.* This group consisted of 13 girls and 45 boys, all aged 9 and 10, from a private school in Omdurman, the capital of Sudan. Unlike the others, these children gave their answers in writing (in Arabic).

### *Treatment of Results*

For the answers to each question, categories (to be described below) were established to which the majority of replies could be assigned. Each answer was assigned to only one category. Answers that did not fall into any of the defined categories are not represented in the tables of results.

## RESULTS

The percentage of each group giving each type of answer to each question is indicated in Table 1. The frequencies for the Lebanese subgroups are shown in Table 2.

The major group contrasts in regard to each question will be discussed in the order in which the questions are listed in the tables. In general, only group differences high in statistical significance are noted. In statements assessing the significance of differences, A-L refers to an American-Lebanese comparison, L-S refers to a Lebanese-Sudanese comparison and A-S to an American-Sudanese comparison.

### *The Mouth*

According to the subjects, the primary uses of the mouth are eating and talking. These two answers make up 100%, 96%, and 92% of the first responses to question 1 for the three groups respectively.

"Talking" as a response to this question is given more frequently by the Americans than by either of the other groups. Table 2 shows that none of the Lebanese subgroups equals the Americans in this regard. Conversely, eating is a response given more often by all segments of the Lebanese sample than by the Americans.

The Sudanese are the only Ss to mention drinking as the first response to the question "What is the mouth for?" Not shown by Table 1, because it deals only with first-mentioned uses, is the fact that among the Sudanese children a very common response to the question, "What is the mouth for?", is "eating and drinking." This joint mention of eating and drinking occurred among 50% of the Sudanese, but among only 3% of the Americans ( $S-A \ p < .001$ ) and none of the Lebanese. The frequent mention of drinking in Sudan is probably related to the high temperatures prevailing there, which require a high water consumption.

It is noteworthy that many possible uses of the mouth were not mentioned by any child in any of the three groups. Many possible replies were absent not only as first responses but failed to appear among subsequent responses as well. Among such absent responses were singing, chewing, spitting, shouting, biting, pointing, smiling, laughing, crying, coughing, sucking, blowing, whistling, praying, and cursing. It seems likely that in societies other than the three which were investigated one or more of these would be common word associations with the mouth.

### *Hands*

For all three groups of children, the uses of the hands are more diversified than the uses of the mouth. The most common responses to the question, "What are hands for?" are to hold things, to grasp, to work, to carry, to touch with, to eat with, and to write with. Only the frequencies of eating, writing, working, and playing are examined below.

Eating as something for which the hands are used is mentioned much more often by the Lebanese and Sudanese (by 23% and 33% of the respective groups) than by the Americans (6%) ( $p$  of A-S and A-L  $< .001$ ). These figures no doubt reflect differences in eating customs among the groups, holding food with the hands being common among some groups of the Lebanese and Sudanese. Among the Lebanese reference to use of the hands in eating is made more often by the government school pupils who in general belong to a lower economic group than by the private school pupils (Table 2).

In regard to writing and working, the response frequencies of the three major groups are almost identical. Play is mentioned most frequently by the Americans, seldom by the Lebanese, not at all by the Sudanese. Other responses are of such a miscellaneous nature as to be difficult to treat in a comparative manner.

### *Mother*

In all groups, providing food and care for children comprise the major functions of a mother in the eyes of the child. Providing food is most often mentioned among the Lebanese. This response is more common among the younger Lebanese children than among the older, but even among the Lebanese 9-10-year-olds food is mentioned more frequently than by the Americans as a whole. The Sudanese children are unique among the groups in specifically mentioning nursing, 19% of their first-mentioned uses of the mother being to nurse a baby. The complete absence of such responses among the American and Lebanese Ss probably reflects a taboo against making references to the breasts in these groups.

Although the general response, "she takes care of us," was common, in all groups there is an al-

TABLE 1  
MAJOR GROUP COMPARISONS

Stimulus Words and Use—Categories	Percentage Frequencies of Responses		
	American	Lebanese	Sudanese
Mouth			
eating	61	74	71
talking	39	22	14
drinking	0	0	7
Hands			
eating	6	23	33
writing	11	10	16
working	12	12	12
playing	4	1	0
Mother			
providing food	16	51	7
providing care	62	44	41
nursing	0	0	19
Father			
working, earning money	55	76	10
buying food	1	16	7
providing care	12	6	16
assisting family	14	3	29
Boy			
going to school	16	37	7
working, helping par- ents	23	24	17
future reference	11	20	12
playing	34	9	0
Girl			
going to school	12	33	9
working, helping par- ents	28	45	34
future reference	12	11	0
playing	27	5	0
Trees			
food	34	69	52
wood	8	8	10
shade	10	2	12
climbing	12	1	0
Wood			
fire	43	33	5
building	34	17	14
furniture	12	33	21
gates	0	0	17
Dogs			
guarding, watching	16	66	62
barking, biting	9	19	0
hunting	4	8	7
pets, playing	49	4	3
Cats			
catching mice	16	63	53
meowing	6	13	2
pets, playing	52	6	7
Birds			
eating	7	37	16
flying, singing	34	36	36
enjoying	33	8	2
Stones			
throwing, hitting	19	13	2
building	38	73	53

TABLE 1—Continued

Stimulus Words and Use—Categories	Percentage Frequencies of Responses		
	American	Lebanese	Sudanese
Sun			
drying	0	9	2
warming	25	28	9
light	53	47	67
Sand			
playing	41	18	2
building	21	58	26
Rain			
growing plants	75	46	41
providing water	19	16	12
Gold			
decorative	16	53	67
economic	61	26	2

most complete absence of reference to *specific* motherly acts other than feeding. No child mentioned sweeping, washing dishes, laundering, ironing, sewing, dressing, washing faces, bathing, shopping, putting to bed, arousing, getting off to school, or the superintending of these activities, although some of these would probably have been mentioned if as many responses as was possible had been obtained to each question. Also unmentioned are helping with lessons, moral instruction, reward and punishment, dispensation of justice within the family, love and affection, story-telling, and parent-child play.

#### Father

If the child's point of view is accepted, the life of a Lebanese father must be a dreary one. He works and earns money in order to support the family. This is the general tenor of all answers. As in the case of the mother, there is no mention of his playing with the child, rewarding or punishing, or of expressing love or hate. So far as the manifest content of the responses are concerned, the Lebanese child seems to have a father who provides for him but who otherwise has little relationship to him. The answers of the American children are similar, but there is a much larger miscellaneous category. In contrast to the Lebanese and the Americans, the Sudanese child mentions the working and earning functions of his father very seldom. Only 10% of the Sudanese answers fall into this category as compared to 55% and 76% of the Americans and Lebanese respectively. Assisting the family and taking care of the family are common answers among the Sudanese but are infrequent in the other two groups.

#### Boy and Girl

"What is a boy for?" "What is a girl for?" Nearly every answer to these questions can be placed in one of four categories: The function of a



TABLE 2  
COMPARISON OF LEBANESE SUBGROUPS

Stimulus	Percentage Frequencies of Responses									
Words and Use—Categories	Moslems	Christians	Armenians	Boys	Girls	Private Schools	Public Schools	5-6 yrs.	7-8 yrs.	9-10 yrs.
Mouth										
eating	81	80	70	79	75	73	75	79	82	70
talking	18	19	29	20	24	27	17	19	17	31
Hands										
eating	28	24	19	21	25	15	30	41	20	10
writing	10	14	10	9	13	15	5	4	13	16
working	22	18	13	11	14	12	12	7	16	27
playing	1	1	4	1	0	1	1	3	2	1
Mother										
providing food	55	42	57	55	47	44	59	65	55	34
providing care	41	53	39	39	49	54	34	28	38	64
Father										
working, earning money	64	77	88	74	78	77	76	70	81	82
buying food	28	11	9	17	15	15	18	23	13	13
providing care	1	4	1	3	3	3	3	2	2	2
assisting family	2	1	0	1	2	2	1	1	1	1
Boy										
going to school	40	37	32	43	30	41	34	40	38	31
working, helping parents	17	18	38	33	26	22	27	16	30	27
future reference	15	28	14	17	23	20	20	8	19	33
playing	8	9	9	9	8	9	9	18	5	3
Girl										
going to school	30	36	32	32	33	35	30	40	33	25
working, helping parents	43	40	49	47	44	41	50	28	44	59
future reference	10	12	10	11	11	12	9	6	13	12
playing	4	5	6	6	5	6	5	11	3	1
Trees										
food	65	70	72	68	70	72	66	63	74	71
wood	10	10	5	8	8	7	9	8	7	12
shade	2	3	2	2	2	2	2	2	2	2
climbing	3	1	1	1	2	1	2	3	1	0
Wood										
fire	40	27	34	33	34	39	28	38	38	24
building	21	22	7	16	17	17	16	21	16	13
furniture	33	48	18	34	32	41	25	16	36	45
Dogs										
guarding, watching	59	68	65	65	63	70	58	37	68	84
barking, biting	17	14	25	16	21	13	24	40	11	5
hunting	11	10	3	9	6	10	6	5	9	8
pets, playing	7	3	1	5	3	3	5	6	4	1
Cats										
catching mice	62	67	54	66	60	71	55	33	69	82
meowing	13	13	13	10	16	9	17	25	10	4
pets, playing	10	7		6	5	4	7	9	4	4
Birds										
eating	40	47	24	36	38	34	40	28	40	45
flying, singing	28	28	50	34	38	35	36	47	35	25
enjoying	11	11	3	6	10	11	5	5	6	13
Stones										
throwing, hitting	18	14	6	13	13	11	14	25	8	5
building	66	71	72	72	67	77	63	50	71	88
roads	4	2	3	3	4	1	5	5	3	2
Sun										
drying	10	11	5	9	9	6	12	10	11	5
warming	32	26	30	23	38	27	30	30	30	29
light	38	46	57	52	42	52	42	34	50	57

TABLE 2—Continued

Stimulus	Percentage Frequencies of Responses									
	Moslems	Christians	Armenians	Boys	Girls	Private Schools	Public Schools	5-6 yrs.	7-8 yrs.	9-10 yrs.
Sand playing building	22 50	20 63	13 60	16 66	21 50	28 51	8 64	30 44	18 58	7 72
Rain growing plants providing water	45 17	50 15	43 17	51 15	42 17	63 13	30 18	25 13	50 19	64 16
Gold decorative economic	60 25	54 28	52 25	39 46	72 12	49 35	62 16	49 17	52 31	65 29

boy (or a girl) is (a) to go to school, (b) or to work, help at home, assist mother and father, do chores and run errands, (c) or to grow up in order to do something in the future (become a man or woman, earn money when he grows up, etc.), (d) or to play.

Among the Lebanese, going to school comprises 37% of the answers referring to the boy and 33% of the answers referring to the girl. Among the American and Sudanese groups school is mentioned much less frequently, although all of the Ss were attending school when questioned. Evidence from other studies, to be reported separately, indicates that schooling has a very high value among Lebanese parents and children.

Working (including the helping of parents) stands relatively high in frequency in all groups. In all groups working is mentioned more often as a function of the girl than as a function of the boy. Preparing for future has frequencies roughly equal in all groups, except that among the Sudanese Ss, most of whom were boys, preparation for the future on the part of girls received no mention.

Play seems to be the outstanding function of American children. It is said to be the purpose of a boy in 34% of the instances and of the girl in 27% of the cases. The corresponding Lebanese frequencies are 9% and 5% ( $p$  of A-L for boys and girls each  $< .001$ ). This finding, too, is congruent with other data to be presented in subsequent reports.

No mention of play occurred in the Sudanese answers to these questions. No supporting data are available, but the rarity of Sudanese references to play strongly suggests either that play is relatively infrequent, or that it is disapproved of, or both, among this group.

### Trees

Sixty-nine per cent of the Lebanese Ss state that trees are to provide food. The high proportion of such answers among the Lebanese is no doubt related to the fact that Lebanon has a great variety

of food-bearing trees, including olives, almonds, hazel nuts, walnuts, dates, oranges, lemons, grapefruit, bananas, cherries, peaches, plums, and apples. Even pine trees are prized for edible nuts. But American children, living in the same environment, mention the food-producing function of trees only half as often (34%) (A-L  $p < .001$ ). This finding demonstrates again that answers are derived from culture rather than from the physical environment.

The shade-producing function of trees is mentioned more often by the Americans and the Sudanese (10% and 12%) than by the Lebanese (2%) ( $p$  of S-L = .01,  $p$  of A-L  $< .001$ ). Climbing trees, which is a play function, has a frequency of 12% among the Americans, 1% among the Lebanese, and zero among the Sudanese. Again, since American and Lebanese Ss were all living in Beirut, it can be assumed that the differences were not produced by differences in natural environment.

### Wood

Wood as a fuel is frequently mentioned by the Americans and the Lebanese, but only seldom by the children of Sudan where heated buildings are unknown. Wood as a material used in building is much more familiar to Americans than to the other two groups. No Lebanese or Sudanese homes are built of wood, this scarce material being reserved chiefly for doors and window frames. Sudanese homes are frequently surrounded by an adobe wall, entrance to which is by a wooden gate. It will be noted that wood used for gates is referred to only by the children of the Sudan.

### Cats and Dogs

In the Near East and in certain other parts of the world cats and dogs are not pets as they are in Europe and America. In the Near East they are treated in much the same way as are other domestic animals whereas in the Western world they have a special status. The Lebanese and Sudanese



answers reflect the fact that if a dog is kept it is for the purpose of guarding the house and warning against intruders. A dog is occasionally used for hunting. In consonance with these adult notions, guarding and hunting make up 62-66% of the Lebanese and Sudanese responses. Among the Lebanese, to play with a dog and to pet a dog are referred to only by the younger children, among whom such answers occur in 4% of the cases. Attitudes towards dogs do not differ as between the Lebanese Moslem and Christian groups (see Table 2).

In contrast, 49% of the American answers indicate that dogs are pets and are to be played with ( $A-L \, p < .001$ ). Among the Americans such answers are as frequent among the older children as among the younger.

The differences among the groups in regard to cats is similar to that prevailing in reference to dogs. For the Lebanese and Sudanese the primary feline purpose is the catching of mice and rats. Occasionally the catching of insects by cats is mentioned. When specific insects are mentioned, they are usually cockroaches. Reference to cats as animals to play with or to pet occurs infrequently in these groups (6-7%) but 52% of the American answers fall into this class ( $p < .001$ ).

### *Birds*

For some children in all groups, birds are to fly, to sing, and to be put in a cage. But 37% of the Lebanese Ss state that birds are to be killed and eaten. This reflects the Lebanese custom of shooting and eating many birds which Americans do not classify as game birds. The reader may note again the prevalence of responses referring to food on the part of the Lebanese. The Sudanese mention birds as food less frequently (16%) than the Lebanese, but more frequently than the Americans (7%) ( $A-L \, p < .001$ ).

### *Stones*

In all groups stones are primarily to build with, 38%, 73% and 53% of the answers, respectively, being in this category. This answer is not surprising in countries in which stone has been the most widely used building material for thousands of years.

"Stone" was included in the stimulus list in part because stones can be used aggressively. In this regard only small differences appear among the groups. Using stones for throwing or hitting makes up only 13% of the Lebanese, 19% of the American, and 2% of the Sudanese responses. That the frequency of this answer is low among the Sudanese seems to be due to the greater age of the Sudanese subjects. Age is a major variable affecting aggressive verbal responses in regard to stones

(see Table 2). Among the Lebanese responses, throwing and hitting make up 25% of the responses of the 5-6 year olds to the word "stone." Beyond this age such responses become infrequent in both Lebanese and American groups, apparently indicating strong disapproval of overt aggression.

### *The Sun*

As might be expected the sun's predominant functions are to provide warmth and light. On this item one of the major group contrasts lies in the fact that the Sudanese Ss referred to the heat-providing function of the sun much less often than the other groups (9% as contrasted to 25-28%) ( $A-S \, p < .001$ ). Apparently, the Sudan is ordinarily so hot that the child thinks the sun's contribution is not useful!

Another service provided by the sun is that of drying clothes. This is mentioned by 9% of the Lebanese, 2% of the Sudanese, and none of the Americans. Presumably, more clothes are dried at home in the former groups. Table 2 indicates that among the Lebanese this answer is more prominent among the very young (10%) than among the oldest group (5%) and more frequent among the poorer than among the more well-to-do (12% and 6%).

### *Sand*

Because of very extensive building operations which are under way in Beirut, construction work is familiar to all Beirut children. The use of sand in building is mentioned by 58% of the Lebanese. American children living in Beirut mention construction much less often (21%) ( $p < .001$ ).

One of the major group differences in regard to sand involves play. Playing in sand is mentioned by 41% of the Americans, 18% of the Lebanese ( $A-L \, p < .001$ ), and 2% of the Sudanese. The generally high rating of the Americans on play responses has been noted in connection with several earlier questions. The low play rating of the Sudanese Ss on this particular question is in part due to their ages, since sand play is uncommon at ages 9 and 10. Among the Lebanese, play as a use for sand has a frequency of 30% at ages 5-6 and of only 7% at ages 9-10. It will be noted that the frequency for American children for ages 5-10 combined is higher than the frequency at ages 5-6 among the Lebanese.

### *Rain*

Rain provided no group differences of special interest. In all groups, the most common use attributed to rain is that of aiding the growth of plants. This response is highest among the Americans. The other two groups gave many more answers which fell into the miscellaneous or unclassified category.

*Gold*

Almost all answers concerning gold fall into two classes: Those referring to decorative uses such as the making of rings, bracelets and earrings, and those referring to economic uses, such as statements to the effect that gold is to sell, or to buy things, or to make money, or to make one rich.

The Sudanese frequently mention decorative uses and seldom mention economic ones. Americans give a reverse picture, economic responses predominating. The Lebanese have intermediate response frequencies. These differences are probably related to variations in the use of gold ornaments among the three groups, and to variations in acquaintance with the monetary value of gold.

In Lebanon, chocolate candies are frequently wrapped in gold foil. As a result, among the younger Lebanese children a not infrequent answer to "What is gold for?" is "To wrap candy in."

*Comparisons of Lebanese Subgroups*

In the foregoing comparisons among the three main groups, reference has been made to Lebanese subgroups only when it clarified the main group contrasts. Let us now examine in more detail the data contained in Table 2.

While Lebanon is noted for its cultural diversity it appears from Table 2 that the children of Moslems, Armenians, and Arab Christians have much in common. The first three columns of Table 2 present data for these groups. In most instances, the percentages for Moslems and Arab Christians, who have lived side by side for centuries, are almost identical. Since the majority of Armenians in Lebanon have immigrated to this country since the Turkish massacres of 1914, they have had a relatively short time in which to become acculturated. Their responses differ somewhat from those of the other Lebanese groups. But they diverge much less from the Moslems and Arab Christians than do the Americans and the Sudanese.

So far as the answers given by the two sexes are concerned, again there is a general similarity at the age levels represented in this study. Only a few significant sex differences appear. In regard to the sun, girls stress warmth more than do boys (38% vs. 23%  $p < .001$ ). Playing in sand is more often mentioned by the girls (21% vs. 16%  $p < .001$ ). The largest difference appears in regard to gold. Decorative uses are given by 72% of the girls and only 39% of the boys ( $p < .001$ ).

Age affects the frequency of responses to nearly every question. Thus, eating as a function of the hands decreases with age, and writing increases. Play in general decreases, whereas work increases. The reader with a special interest in age differences will wish to examine Table 2 in some detail. It will

not be feasible to comment on age changes in respect to each item. Concepts of use strongly reflect developmental age. This fact underlines the importance of equating for age any groups which are to be compared. The reader should bear in mind that the American and the Lebanese groups were equated for age composition. The fact that the Sudanese group consisted entirely of 9- and 10-year olds has been taken into account in making all comparisons.

*Hypotheses Concerning the Lebanese Child*

The data have been seen to reveal a considerable number of differences among the three major groups of Ss. It remains to attempt to characterize in a general way the answers of each group. Such a general characterization is warranted only for the Lebanese and the American groups in view of the small size of the Sudanese sample and the fact that it represented only one age level.

It has already been observed that the Lebanese answers seem to show a considerable concern with food. Common answers include the following: mother feeds us, father provides food, the mouth and the hands are used for eating, trees produce food, and birds are to be eaten. The answers taken by themselves do not show why food is so important. There is no reason to believe that the Ss suffered from malnutrition. Actually, the majority of Lebanese children appear to be very well fed. Observations other than those presented here support the tentative suggestion that food preparation and eating have for the Lebanese a pleasure function that is comparable to the positive valuation given by some societies to other activities such as music, conversation, or visual esthetic experiences.

In the second place, the Lebanese responses seem to indicate that industriousness is highly valued. Note the following common answers: father and mother work, girls and boys work, the hands are for work, dogs work by guarding the house, cats work by catching mice and rats. Trees and birds provide food, which may be thought of as a work function. Stones and sand supply building materials. Much of the world seems to be oriented toward utilitarian ends.

Concern for language facility and scholastic performances seems evident, particularly among the private school pupils. The mouth is for talking, the hands for writing, and one of



the chief functions of both boys and girls is to go to school.

Few other emphases are observable in the Lebanese answers. There is some mention of play in connection with questions about boys and girls, and the question about sand, but little or no mention of it in reference to parents, hands and mouth, or animals. There is no mention of art, of music, of athletics, or of sensory pleasures other than eating. There is little indication of valuation of emotional life or of group cohesion. There are no references to competitiveness in areas other than school achievement. There is nothing to suggest that religious observances or religious affiliations are highly significant.

### *Hypotheses Concerning the American Child*

The present data, it is important to remember, are restricted to American children in Lebanon, and more specifically, to those attending a specific progressive school. The parents are nearly all college graduates, and most seem to have accepted the doctrine of permissiveness. While many groups of children living in America probably would answer as do the American group in Lebanon, not all groups of children in America would do so.

The Beirut colony of American children (at ages 5-10) may be characterized, from their answers, as being greatly concerned with play. To a much greater extent than among the Lebanese Ss, the hands are for play, boys and girls have play as their purpose, trees are for shade and climbing, dogs and cats serve as pets and playmates, birds are to be enjoyed, and sand serves a play function. In contrast to the seriousness of the Lebanese children, life for the American children in Beirut would seem to constitute a perpetual vacation.

Positive mention of play is associated with low frequencies of reference to work and to duties. Going to school, and helping parents are reported as functions of boys and girls much less often by Americans than by the Lebanese. The function of the father as a bread-winner is less frequently cited by the American group. As noted previously, dogs and cats are for fun and are seldom thought of as utilitarian animals. Similarly, the use of stones and sand as building materials is stressed much less by the American group. Apparently the American child sees the world more in terms of

pleasure and less in terms of duty and work than does the Lebanese child.

This difference appears not to be due to any appreciable extent to differences in social class. If the American group is compared with the Lebanese private school group (Table 2) rather than to the entire Lebanese sample the differences remain substantially unchanged. It is also true that Americans diverge as much from Lebanese Christians as from Lebanese Moslems, these two groups scoring very similar percentages in the response categories to every question (see Table 2).

The categories listed in Table 1 include only the most common answers. The responses that are not represented in this Table were given by only a few children. They may therefore reasonably be called individual or unconventional answers, and may represent individuality or originality. If one adds for each question the percentages of Americans and Lebanese giving answers in the categories of Table 1, one finds that to fourteen of the sixteen questions the Lebanese gave a larger proportion of common, or stereotyped answers. The American answers, in general, show more variety. This finding may represent the result of permissiveness, it may point to more varied experiences, or it may be the result of direct encouragement of individual interests and modes of thought. This difference seems to the investigator to be among the more interesting ones revealed by the study.

### SUMMARY

A technique for the study of children's concepts of the uses of common objects has been presented which is believed to be universally applicable. Its use has been illustrated with American and Lebanese groups and to a lesser extent with Sudanese Ss. The study has demonstrated that in regard to universally present objects, such as hands and mouth, dogs and cats, sun, sand and stone, children of different cultures differ in their ideas as early as five years of age. It has also demonstrated that the test responses enable one to formulate hypotheses concerning the interests, values, and activities which are being developed in different groups of children. The method is believed to be a useful addition to the research techniques available for the study of cultural differences in child development.

Received July 30, 1956.

# AUTHORITARIANISM, INTOLERANCE OF AMBIGUITY, AND RIGIDITY UNDER EGO- AND TASK-INVOLVING CONDITIONS<sup>1</sup>

THEODORE MILLON

*Lehigh University*

**I**N SPITE of extensive research dealing with the connection between measured authoritarian attitudes and behaviors referred to in the literature as rigidity and intolerance of ambiguity, the existence of a true covariance is still in doubt and the antecedent conditions necessary for such a covariance remain unclear. Early research by Rokeach (9) and Frenkel-Brunswik (1, 7) obtained positive relationships between California E- and F-scale scores and intellectual and perceptual measures of rigidity and intolerance of ambiguity. Numerous subsequent studies have failed to substantiate their findings. The view that an atmosphere of ego-involvement or personal stress may be a necessary condition for obtaining such relationships has been recently advanced (3). Unfortunately, experiments designed to test this view have merely contributed additional conflicting evidence (5, 6). The problem has been complicated further by suggestions that induced ego-stress in itself is sufficient to produce rigid and ambiguity-intolerant behavior in experimental subjects (Ss), independently of supposedly predisposing personality traits. Research regarding this point has likewise been equivocal (3, 4, 6, 10).

The number and complexity of contradictory experiments in this area make it impossible to distinguish common factors differentiating studies that obtain a positive relationship between the California scales and rigidity and ambiguity-intolerance from those in which such a relationship has not been found. Some of the difficulty can be attributed to differences in experimental procedures used to induce ego-stress and perhaps even more from a failure to evaluate methods for operationally defining rigidity and intolerance of ambiguity. Ego-involvement has usually been adminis-

tered to large groups with no attempt made to evaluate objectively the effectiveness of experimental instructions. Tasks selected to measure rigidity and intolerance of ambiguity have been many, most having little more in common than the name. Few of the methods have experimentally differentiated the behavioral effects of ambiguity-intolerance from those of rigidity, and the majority appear not to measure variables relevant to the California theory of authoritarianism (1).

The present study was designed to introduce some clarity by: (a) administering ego and task orientations individually, and running postexperimental checks on their effectiveness, and (b) operationally defining and distinguishing between intolerance of ambiguity and rigidity in a manner appropriate to theoretically posited authoritarian characteristics.

In terms of the experimental tasks constructed for the present research, intolerance of ambiguity refers to the tendency to respond to previously unexperienced and unstructured stimuli with a subjectively structured response set. Rigidity refers to the tendency to resist modifying an established response set when the stimulus conditions under which it was established have been changed. To assure that resistance to change is attributable to personality factors and not to initial differences in the habit strength of sets, experimental control of set strength among Ss is necessary. This control is also essential in order to differentiate the behavioral effects of ambiguity-intolerance and rigidity. For example, if Ss in an unstructured stimulus situation are given an *equal* number of trials to establish a response set, Ss high in ambiguity-intolerance may be expected to establish a stronger set than Ss low in ambiguity-intolerance because of their tendency to acquire it earlier and apply it more consistently. Under these conditions, it is psychologically meaningful to speak of rigidity, as gauged by resistance to extinction, only when differences in set strength among Ss

<sup>1</sup> This paper is adapted from a dissertation submitted in candidacy for the Ph.D. degree at the University of Connecticut. The writer is indebted to Maurice L. Farber, thesis advisor, and W. L. Jenkins for assistance in the present formulation. A report of this study was given at the APA meeting, 1954.



are equalized. If rigidity is to serve as a motivational concept, it must contribute to resistance to extinction above and beyond simple habit strength. Unfortunately, many of the operational measures of rigidity currently in vogue fail to assess and control for differences in initial set strength.

In light of these considerations, the following questions were posed for experimental analysis: 1. Do authoritarians subjectively establish a set or norm in a previously unexperienced and unstructured situation more rapidly than equalitarians? 2. *After establishing an equally strong response set*, are authoritarians more resistant to modifying their set than equalitarians when stimulus conditions are changed? The experiment was further designed to answer these questions: 3. Does induced ego-involvement increase relationships among measures of authoritarianism, ambiguity-intolerance, and rigidity? 4. Does ego-involvement in itself produce greater rigidity and intolerance of ambiguity than nonego-involvement?

### METHOD

The Ss whose scores on the standard 30-item Form 40-45 California F scale (1) were known, were asked to participate in a one-session laboratory experiment involving two series of judgments of autokinetic movement. The autokinetic phenomenon satisfied the requirements of unstructuredness and freedom from related prior experiences. The fact that positive results emerged from the only previous study relating the California scales to autokinetic judgments (2) gave further evidence of its appropriateness as a research instrument.

Series 1 consisted of judgments of the direction and distance of movement of a one-light autokinetic stimulus and was designed to measure the speed of norm or set formation; scores were represented by the number of stimulus presentations necessary before S attained a standard experimental criterion of establishing a response set. The criterion required that 10 out of 13 consecutive judgments be reported as moving in only one of four primary directions (up, down, left, right) with a magnitude varying not more than one inch from one another. On the basis of previous research, the assumption is warranted that barring significant changes in the stimulus field, Ss would retain this response set as a frame of reference for subsequent judgments. Series 1 was ended for each S whenever a sequence of his judgments met the criterion. Although the number of trials to reach the criterion varied among Ss, the same degree of consistency in judgments was required of all Ss as the basis for terminating Series 1. Set strength among Ss was thus operationally equated at the conclusion of Series 1, and since the second series followed immediately, Ss began Series 2 with equal

TABLE 1  
F-SCALE GROUPS SELECTED FOR EXPERIMENTATION

Group	N	Approximate F-Scale Centiles	Range of F-Scale Scores	Group F-Scale Mean
L <sub>10</sub>	14	0-10	1.6-2.9	2.52
L <sub>20</sub>	15	11-20	3.0-3.1	3.05
M	13	45-55	3.5-3.8	3.67
H <sub>20</sub>	14	80-89	4.3-4.5	4.35
H <sub>10</sub>	13	90-99	4.6-5.2	4.84

habit strengths in their set for the one-light situation. During Series 2, a second light was added to the stimulus field in conjunction with the first light. The Ss were simply told at this point to continue judging the distance moved by the first light. Preliminary experiments (8) had found that the simultaneous viewing of two lights typically curtailed movement from that perceived under one-light conditions. Series 2 consisted of 50 judgments and was designed to determine the extent to which S shifted from his established one-light norm or set when viewing the two-light stimulus.

Two operational measures were obtained in the autokinetic sequence: intolerance of ambiguity by the speed the norm was formed, and rigidity by the relative distance toward zero S shifted from his one-light norm during the two-light series.

Approximately half of the Ss received instructions designed to create an ego-involving atmosphere, i.e., the impression that the tasks reflected their abilities and intelligence; the other half received instructions to create a task-involving orientation, i.e., the impression that the experiment was designed to try out apparatus for subsequent research on vision. A posttest questionnaire was administered to check on the effectiveness of the induced orientations at the close of the laboratory session. Detailed descriptions of the procedure and questionnaire may be found elsewhere (8).

### Subjects

The Ss were 69 students enrolled in an introductory psychology course. They were assigned at random to one of the two experimental orientations from five divisions of a California F-scale distribution of 162 students. Table 1 presents the centile and F-scale score range for these divisions. Nine Ss failed to establish a norm within the experimentally predetermined maximum of 150 Series 1 stimulus presentations. Norm formation scores for these Ss were arbitrarily set at the maximum of 150. Of the remaining 60 Ss who had met the criterion of norm formation prior to the 150th trial, two subsequently complained of illness and asked to be excused during Series 2.

### RESULTS

An answer to the question of whether authoritarians are more ambiguity-intolerant and rigid than nonauthoritarians can be gathered from Tables 2 and 3. The analysis of scores on trials to form a norm, contained in Table

TABLE 2

ANALYSIS OF VARIANCE OF TRIALS TO FORM A NORM  
(Scores transformed to reciprocals)

Source	df	Mean Square	F	p
Orientation	1	2.00	.61	n.s.
Authoritarianism	4	33.95	10.77	.001
Interaction	4	1.05	.32	n.s.
Within	59	3.30		

TABLE 3

ANALYSIS OF VARIANCE OF PERCENTAGE OF SHIFT FROM NORM

Source	df	Mean Square	F	p
Orientation	1	13,922.5	26.05	.001
Authoritarianism	4	1,293.2	2.42	.06
Interaction	4	657.9	1.23	n.s.
Within	48	534.3		

2, indicates a clear-cut relationship ( $F = 10.77$ ,  $p < .001$ ) between intolerance of ambiguity and authoritarianism. The relationship between authoritarianism and percentage shift from norm, noted in Table 3, yielded an  $F$  of 2.42 ( $p < .06$ ). Although suggestive, this finding does not offer clear support for a correspondence between the California scale and rigidity.

The second major question of the study was whether induced ego-involvement would increase relationships among the measures. Figure 1 illustrates the mean number of trials to norm formation for each of the five F-scale groups under both ego and task orientations. The relationship between rigidity and F-scale scores in terms of percentage of shift is portrayed in Figure 2. It can be seen from this figure that the relationship between F-scale scores and rigidity is more marked in the ego-oriented group. Task-oriented authoritarians, for example, evidence little more rigidity than task-oriented equalitarians; the difference in the ego-oriented condition, in contrast, is quite pronounced. The interaction  $F$ s of .32 and 1.23 noted in Tables 2 and 3 indicate, however, that orientation factors have not been demonstrated to play a statistically significant role in obtaining the generally high relationships found between authoritarianism, ambiguity-intolerance, and rigidity.

The effect of induced ego-involvement in producing increased ambiguity-intolerance and rigidity among experimental Ss is also illustrated in Figures 1 and 2. Although the 9.8 mean difference between ego- and task-oriented groups in trials to norm is suggestive, it is clear from Table 1 that the difference ( $F = .61$ ) does not attain statistical significance. In contrast, a 21.1 per cent greater resistance to shift on the part of the ego-oriented group is significantly greater ( $F = 26.05$ ) than that shown by the task-oriented Ss. This difference is independent of F-scale scores. Figure 2 clearly shows that all of the ego-oriented F-scale groups evidenced considerably more rigidity than comparable groups receiving the task orientation.

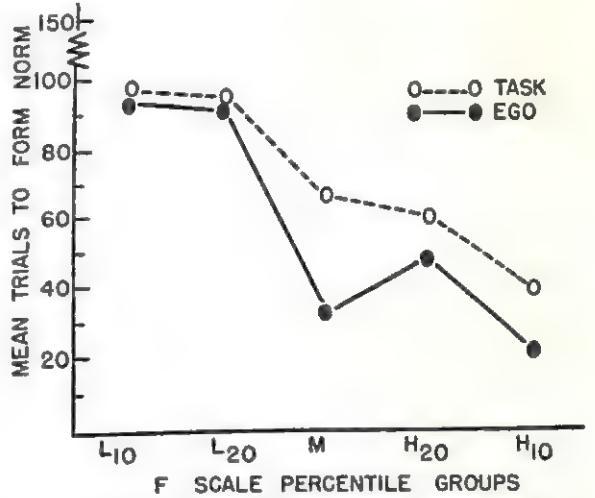


FIG. 1. MEAN NUMBER OF TRIALS TO FORM NORM UNDER EGO- AND TASK-ORIENTING CONDITIONS

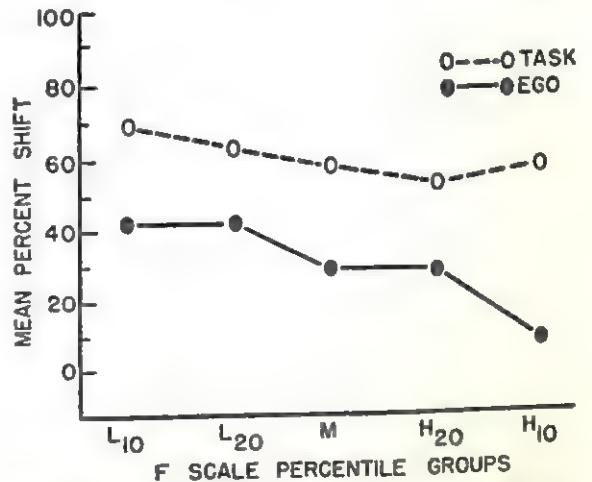


FIG. 2. MEAN PERCENT OF SHIFT TOWARD ZERO FROM ONE-LIGHT NORM UNDER EGO- AND TASK-ORIENTING CONDITIONS



The introspective self-involvement ratings of the posttest questionnaire established a significant difference in the direction expected ( $F = 11.56$ ) between ego- and task-oriented groups. The questionnaire indicated further, that task-oriented Ss showed an appreciable degree of concern regarding their personal performances in spite of the experimental instructions used. No systematic difference was found among the F-scale groups in receptivity to the two testing orientations.

### DISCUSSION

The fact that authoritarians formed norms in the unstructured autokinetic situation more rapidly than nonauthoritarians under both ego- and task-orienting conditions clearly supports Frenkel-Brunswik's position (1, 7) concerning the characteristic ambiguity-intolerance of authoritarians. The general finding that authoritarians modified their established norms to an appreciably lesser degree than nonauthoritarians, offers additional support for the association between rigidity and authoritarianism asserted by the California group. There appears, however, to be a limitation to this covariance. Although authoritarians manifested greater rigidity than nonauthoritarians in the ego-orienting atmosphere, no such difference was evidenced in the task group. Though perhaps a favored mechanism it would appear that rigidity in the authoritarian is partially contingent upon variables associated with ego-involvement. What factors mediate this relationship is problematic.

Although induced ego-involvement cannot be considered to have played a statistically significant part in contributing to relationships found between authoritarianism, ambiguity-intolerance, and rigidity, it is of interest to note that had the experiment been conducted under task-orienting conditions only, the relationship between rigidity and the F scale would have been reported as not significant ( $F = .462$ ). In contrast, had the experiment been run under ego-orienting conditions, the results would have been considered significant ( $F = 3.40$ ;  $p < .03$ ). Although minimal faith should be put in an hypothesis when the statistic called for by the experimental design turns up a nonsignificant effect, it seems wise, in light of the above facts, to reconsider the possible contribution of ego-involvement in

the rigidity-authoritarian relationship. This caution appears particularly justified since the appreciable degree of ego-involvement found among our task-oriented Ss made it impossible to evaluate experimentally the full effects of the ego-involvement variable. Had the difference between the two orientations been more marked, its contribution to authoritarian rigidity may have been more clearly established. In any event, further study of ego-involvement in this context would appear warranted.

A major question underlying the present study was whether personal stress or ego-involvement would increase the general level of ambiguity-intolerant and rigid behavior among Ss regardless of their standing on the F scale. Although the influence is a dubious one as regards intolerance of ambiguity, there can be no doubt as to its effects on behavioral rigidity. Whether ego-involvement was in itself sufficient to bring about the increase cannot be readily determined. The following speculations regarding the connection may nevertheless be offered: Ss who were ego-involved with respect to their one-light norm were motivated to ward off stimuli that conflicted with or questioned the adequacy of this norm. Because of its tendency to curtail the impression of movement, the second light had the effect of challenging the S's established one-light norm. Inasmuch as past judgments presumably reflected intelligence and ability, Ss, in an attempt to safeguard self-esteem and reassure themselves of the adequacy of their one-light performance, warded off the visual influence of the second light. Task-oriented Ss, not feeling a personal identification with their performances in the one-light situation had little or no reason to retain or defend their norm. The thesis suggested then is that ego-involvement in the context of the laboratory tasks of this experiment served to motivate Ss to ignore or reject what was antithetical to their first-formed impressions. It would therefore seem plausible to hypothesize that the effect of a norm-changing stimulus would vary with its acceptability to an individual and that the more ego-involved an individual is with respect to his norm, the more intense will be the operation of his defenses against stimuli that challenge it.

The above interpretation of rigidity as a defensive reaction to the threat posed by the

two-light stimulus adds an additional interpretive problem. Since authoritarians evidenced considerably greater rigidity than nonauthoritarians in the ego-orienting condition, the issue arises as to whether this difference might be better attributed to greater ego-threat experienced by authoritarians than to greater rigidity manifested by them in response to equivalent degrees of threat. Because of its highly structured and limited nature, the posttest questionnaire offered no effective clues as to the validity of this hypothesis. If authoritarians were in fact more threatened than nonauthoritarians when the second light was presented, their greater resistance to change could be considered commensurate with the degree of threat they experienced. If this speculation is valid, authoritarians would be characterized better by their susceptibility to intense threat when their personal norms are questioned than by their "generalized mental rigidity."

Space does not justify a detailed review of how results of the present study agree or disagree with earlier research. Obviously, there are striking differences between experimental techniques used for measuring intolerance of ambiguity and rigidity. The fact that the present ego- and task-orienting instructions were given person-to-person rather than in a group setting, as is more usual, is a further obstacle to comparison. Until additional unequivocal data are available to reconcile contradictory evidence from previous studies, it seems wisest to view new positive results concerning the relationship of rigidity and tolerance of ambiguity to authoritarianism with caution. Clearly, these behaviors are influenced not only by personality factors, but also by the tasks utilized for measurement and the situational atmosphere within which the S must respond to them.

#### SUMMARY

Individual laboratory sessions involving judgments of autokinetic stimuli were conducted with 69 college students whose standings on the California F scale were known; approximately half received ego-orienting instructions, half received a task orientation. Two behavioral scores were obtained during the experimental sequence: (a) the number of trials to form a norm while estimating direction and distance of a one-light autokinetic stimulus, as a measure of intolerance of ambiguity,

and (b) the percentage of appropriate shift from the S's established one-light norm upon the introduction of second light, as a measure of rigidity.

Analyses of variance of these measures produced the following findings:

1. (a) An  $F$  beyond the .01 level of confidence between authoritarianism and intolerance of ambiguity.

(b) An  $F$  attaining the .06 level between authoritarianism and rigidity. The relationship was clearly significant in the ego-oriented group, but not in the task-oriented group.

2. An  $F$  beyond the .01 level of confidence signifying greater rigidity in the ego- than the task-oriented group.

3. Ego-involvement did not significantly increase relationships among the three experimental measures, nor did it produce a general increase in ambiguity-intolerance.

The results of the study were felt to be consonant with positions taken by Frenkel-Brunswik regarding the typical rigid and ambiguity-intolerant behavior of authoritarians. Limitations to this relationship were noted and alternate hypotheses suggested.

#### REFERENCES

1. ADORNO, T. W. *et al.* *The authoritarian personality*. New York: Harper, 1950.
2. BLOCK, J., & BLOCK, J. An investigation of the relationship between intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1951, **19**, 303-312.
3. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469-476.
4. COWEN, E. L. The influence of varying degrees of psychological stress on problem-solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512-519.
5. DAVIDS, A. The influence of ego-involvement on relations between authoritarianism and intolerance of ambiguity. *J. consult. Psychol.*, 1956, **20**, 179-184.
6. FRENCH, E. G. Interrelation among some measures of rigidity under stress and nonstress conditions. *J. abnorm. soc. Psychol.*, 1955, **51**, 114-118.
7. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108-143.
8. MILLON, T. The role of situational variables in perceptual behaviors characteristic of authoritarians. Unpublished doctor's dissertation, Univer. of Connecticut, 1954.
9. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259-278.
10. SMOCK, C. D. The influence of stress on "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1955, **50**, 177-182.

Received July 31, 1956.



# COGNITIVE COMPLEXITY AND ASSIMILATIVE PROJECTION IN ATTITUDE CHANGE

RICHARD M. LUNDY AND LEONARD BERKOWITZ

*University of Wisconsin*

**A**MONG the factors that have been investigated in studies of attitude change (e.g., 4, 5) are the personality of the person being influenced, and the source of the attitude change material. The present study was designed to test hypotheses in each of these general areas.

According to the first hypothesis, attitude change is more likely to occur in "cognitively complex" persons than in "cognitively simple" persons. The concept of cognitive complexity-simplicity stems from the theoretical orientation of Kelly (6), which assumes that each individual possesses a system of constructs for understanding and predicting his world. In comparison with a simple system, a complex system differentiates more sharply among ideas. It distinguishes among a greater number of different aspects of people or situations, employing more functional constructs and showing more variety in the relationships among them.

A basic assumption in the present study is that the ideas expressed in a communication have meaning for its recipient and can therefore be considered by him only if they are relevant to his existing personal constructs. The complex system, having more functional constructs, furnishes greater opportunity, statistically, for a new idea to be considered. Once considered, the new idea may be integrated into the person's construct system to result in attitude change. In contrast, the cognitively simple system, equipped with fewer functional constructs, allows fewer opportunities for new ideas to come in relation to a personal construct and to be integrated into it. *New words may be manipulated, but unless the new idea is related to a relevant personal construct it can effect no change in the person's construct system and, in a sense, has no meaning for him.*

This formulation resembles the Gestalt hypothesis that attitude changes normally occur because the "material to be judged is seen in a new light and has consequently changed its meaning" (7, p. 19). Both views emphasize

the importance of the meaning of the communication to the recipient. The present formulation sets up special conditions for "change in meaning" by stating that the new material must have relevance for an individual's personal constructs before it can be "seen in a new light."

The remaining hypotheses concern assimilative projection or assumed similarity (3, 8) as relating the source of attitude change material to the person being influenced. Assimilative projection refers to the degree to which a person perceives other persons (the communication source in this case) as similar or dissimilar to himself. If a person perceives the source of a new idea as similar to himself, he is likely to see the idea also as similar, in some ways, to his own ideas, and thus consider it "correct." He should therefore be more susceptible to influence by a particular source group if he perceives that source group as similar to himself. In the present study, two source groups are used—authority and peers. Hypothesis II states: Persons who are susceptible to peer influence are more likely to perceive peers as similar to self than persons who (a) are negative to peer influence, or (b) are susceptible to authority influence. Hypothesis III states: Persons who are susceptible to authority influence are more likely to perceive authority figures as similar to self than persons who (a) are negative to authority influence, or (b) are susceptible to peer influence.

## METHOD

### *Attitude Change Measure*

During another investigation by the authors (1) a 132-item attitude questionnaire was administered to students (male and female) in two intermediate psychology courses. Scattered throughout the questionnaire were items from three modified Thurstone scales: attitudes toward the movies, tariff, and the punishment of criminals. The students were told the questionnaire involved an investigation into the "psychology of normal people" concerned with their attitudes on "every-day matters."

One month after the original testing session, four alternate forms of the second questionnaire were administered to the two classes. These questionnaires

were introduced with the following orally-delivered statement:

Many investigations into the psychology of attitudes have dealt with issues generally given little thought or attention by people. We want to study attitudes that have been given some thought. In order to stimulate your thinking on these issues we have provided you with the opinions of other people. For each of the three attitude issues, first read the opinions expressed by these other people, then indicate your opinion on each item.

Half of the questionnaires passed out to the classes presented the opinions supposedly expressed by peers with this introduction:

A national public opinion organization recently conducted a survey of the opinions of college students in the universities of the Big Ten... The survey showed a fairly high degree of consensus among the college students on several issues. On the following pages we have presented the attitudes of the great majority of college students.

The remaining half of the questionnaires contained opinion statements supposedly expressed by generals. The students were told in these questionnaires:

A writer for a national magazine recently interviewed Generals Bradley, Clark, and Taylor on certain issues of national interest. The generals were in complete agreement on these issues. On the following pages we have presented their attitudes.

In contrast to the questionnaire administered in the first session, the items pertaining to a given issue, i.e., either movies, tariffs, or the punishment of criminals, were presented together on a single page. Before each page containing the given attitude scale was a page describing the opinions of either the peers or the generals on this issue. To illustrate, the items in the attitudes-toward-movies scale were preceded in some of the booklets by a page containing this statement: "The great majority of the college students expressed favorable attitudes toward movies. They felt that movies often were very fine entertainment and frequently contributed to the education of the moviegoer." For each attitude scale, half of the booklets contained favorable opinions by either the peers or generals while the remaining booklets contained unfavorable statements. The same opinion rationale was used for both the peers and generals. After reading the introductory statement to each issue, the subjects (Ss) indicated their opinions on the given attitude items by marking an IBM sheet.

### *Cognitive Complexity-Simplicity*

Cognitive complexity-simplicity was measured by the Role Construct Repertory Test (RCRT), an instrument suggested by Kelly and used by Bieri (2) in his study on cognitive complexity. The test consists of a grid across the top of which S writes the names of 22 people personally known to him, who fit in prescribed role categories. The S is asked to consider three of these people at a time and to decide in what important way two of them are alike and different from the third. Different groups of three are considered successively by S until he has listed along the side of the grid a certain number of personal constructs derived from the original list of people. For each construct thus derived, the S

then considers all the rest of the people in his list and places a check mark under the names of those people perceived as similar to the original two in terms of the construct. In this way the grid is formed with checks or voids which indicate the particular manner in which the S perceives similarities and differences within a group of people on the basis of his personal constructs.

By comparing each construct row of check marks with each other construct row, one can determine their degree of similarity. If two rows have check marks under the same names, the constructs are functionally equivalent. When many of the construct rows for an S are alike, he is considered as cognitively simple. If he produces few rows that are identical or similar to any other row, he is regarded as having high cognitive complexity. The scores for cognitive complexity-simplicity were obtained by using the first 10 rows and first 10 columns in the standard RCRT.<sup>1</sup> Each row was matched with each other row. When a matching was identical, a score of three was given; when a matching was identical except for one, a score of two was given; and when except for two, a score of one. A high score indicates cognitive simplicity; a low score, cognitive complexity.

### *Assimilative Projection*

One of the persons which the S is instructed to use in the RCRT is himself. After the grid is completed, the column of checks and voids under the S's name indicates how he perceives himself in regard to his own constructs. By comparing this self perception with his perception of other people as indicated by his test responses, his degree of perceived similarity, i.e., assimilative projection, can be ascertained. To find assimilative projection with peers, four figures were selected from the first ten on the grid to be matched with the self: (a) a brother, or brother-like person, (b) a sister, or sister-like person, (c) the present closest opposite-sex friend, and (d) the present closest same-sex friend. The assimilative projection score for peers was made up of the number of times a check or void under the self matched a check or void under each of the peers considered separately.<sup>2</sup>

The assimilative score for authority figures was obtained the same way, using: (a) the mother figure, (b) the father figure, (c) the minister, priest, or rabbi with whom personal feelings about religion might be discussed, (d) the physician, (e) the teacher who influenced S most when in the teens, (f) the S's employer, supervisor, or officer during a period of great stress. In scaling the items, it was found that mother figure, father figure, and employer figure, individually, correlated very little with the total score. In the subsequent analysis of data, therefore, only the religious person, the physician, and the teacher were used in determining the score on similarity to authority. The latter two were not among the figures included in determining the cognitive complexity score.

<sup>1</sup>To facilitate scoring, the first ten rows and columns were used instead of the complete grid.

<sup>2</sup>An analysis of the data revealed a slight negative relationship between assimilative projection on peers and cognitive complexity. Since both these variables are positively related to attitude change, this slight negative relationship indicates that the tests of the two hypotheses are independent of each other.



TABLE 1  
COGNITIVE COMPLEXITY-SIMPLICITY IN RELATION  
TO SOURCE OF INFLUENCE AND TYPE OF  
ATTITUDE CHANGE

Type of Attitude Change	Mean Cognitive Complexity-Simplicity*	
	Generals (N) as source	Peers (N) as source
Little change	(9) 16.44	(6) 35.50
Variable change	(5) 16.20	(9) 14.78
Susceptible	(7) 13.43	(9) 14.89
Negative change	(6) 8.83	(12) 12.58

Analysis of Variance				
Source of Variation	DF	MS	F	p
Source of influence	1	200.16	—	—
Type of attitude change	3	570.56	5.0	.05
Interaction	3	297.13	2.6	—
Within groups	55	114.19		

\* A high score represents greater simplicity.

### Analysis of Data

Change scores from the first to the second testing were determined for each of the three scales. These raw scores were then converted into standard scores with a mean change score of 20 for each scale, and the three scale scores added to obtain a total standard susceptibility score. Since the mean change score over the three scales was 60, an S scoring higher than 60 was above the average in over-all susceptibility to change. Two groups of Ss were formed using only these total standard scores: (a) *susceptible* group, Ss having standard scores of 67 or more; (b) *negative change* group, Ss having total standard scores of 54 or less.<sup>3</sup>

Ss whose total scores ranged close to the mean change were divided into two groups according to the variability of their reactions among the three scales. On the basis of a variability score, yielded by the sum of the differences between the S's standard score on each scale and the corresponding mean standard score, the third and fourth groups were established as follows: (c) *little change* group, Ss having total standard scores of 55 to 66 and variability scores of 7 or less; (d) *variable change* group, Ss having total standard scores of 55 to 66 and variability scores of 8 or more. The scores defining the

<sup>3</sup>The standard scores were combined in order to obtain a measure of the individuals' general reactions to the external communication. The Ss tended to be slightly influenced in the direction of the communications. The raw score change means on the three attitude scales were 0.0, 1.0, 2.0, and the SDs were 3.7, 3.5, and 6.0, respectively.

The *susceptible* Ss generally changed in the direction of the communication. Twelve of the 16 Ss changed toward the communication on all three scales; the other four were successfully influenced on two out of three scales. The *negative change* Ss tended to change in the direction opposite to that advocated by the communication. Sixteen of the 18 Ss changed away from the communication on two or three of the three scales.

four groups were selected so that the Ns in the groups would be nearly equal.

To study the personality characteristics associated with the four kinds of reactions to the communication, the four attitude change groups were used in setting up analysis of variance designs with cognitive complexity-simplicity, perceived similarity to peers, and perceived similarity to authority as dependent variables in the analyses. If attitude change had been taken as the dependent variable for purposes of the analysis, these kinds of reactions to outside influence could not readily have been differentiated. An N of 63 in the complexity analysis was reduced to 61 in the assumed similarity measures because two Ss had not recorded any marks in the self column of the RCRT. Because of disproportional cell frequencies, the analysis of variance procedure suggested by Snedecor (9) was employed.

TABLE 2  
ASSIMILATIVE PROJECTION ON PEERS IN RELATION  
TO SOURCE OF INFLUENCE AND TYPE OF  
ATTITUDE CHANGE

Type of Attitude Change	Mean Assimilative Projection on Peers	
	Generals (N) as source	Peers (N) as source
Little change	(9) 19.56	(6) 24.50
Variable change	(5) 20.80	(9) 21.89
Susceptible	(6) 20.50	(8) 23.75
Negative change	(6) 24.17	(12) 19.50

Analysis of Variance				
Source of Variation	DF	MS	F	p
Source of influence	1	11.18	—	—
Type of attitude change	3	5.21	—	—
Interaction	3	67.28	4.12	.05
Within groups	53	16.34		

TABLE 3  
ASSIMILATIVE PROJECTION ON AUTHORITY FIGURES  
IN RELATION TO SOURCE OF INFLUENCE AND  
TYPE OF ATTITUDE CHANGE

Type of Attitude Change	Mean Assimilative Projection on Authority Figures	
	Generals (N) as source	Peers (N) as source
Little change	(9) 14.67	(6) 16.67
Variable change	(5) 13.60	(9) 15.67
Susceptibility	(6) 17.33	(8) 14.75
Negative change	(6) 12.67	(12) 12.92

Analysis of Variance				
Source of Variation	DF	MS	F	p
Source of influence	1	.08	—	—
Type of attitude change	3	31.11	2.47	—
Interaction	3	16.23	1.29	—
Within	53	12.60		

## RESULTS AND DISCUSSION

Hypothesis I states that attitude change is less likely to occur in "cognitively simple" persons. Since the "little change" group contains persons who maintain their attitudes in the face of the outside influence, this group should have the greatest cognitive simplicity. The results of the analysis of variance given in Table 1 indicate that there are significant differences among the attitude change conditions when the communication source is not considered. The little change group has the highest cognitive simplicity means, tending to support the hypothesis.

The mean cognitive simplicity in the "little change to peers" condition is significantly greater ( $p < .05$ ) than the mean in the "little change to generals" condition, suggesting that the hypothesized relationship between little change in the attitude scale and cognitive simplicity is most pronounced when the source of the attitude influence is the Ss' peer group. It may be that a "cognitively simple" person is not as resistant to change when the source is authority figures because of a need to conform to these authority figures. Interestingly, the groups with the highest cognitive complexity are those negative in their reactions to both peer and authority persuasion.

Hypothesis II states that persons who are susceptible to peer influence are more likely to perceive peers as similar to self than persons who are negative to peer influence or are susceptible to authority influence. This hypothesis is supported by the significant interaction ( $p < .05$ ) in the analysis of variance presented in Table 2. The Ss in the "susceptible to peers" condition perceive themselves more similar to peers than the Ss in the "negative to peers" condition ( $p < .05$ ), while the direction of the differences between the susceptible and negative conditions is reversed when generals are the source of influence. Although this latter difference does not attain significance, it may be that Ss who perceive themselves similar to peers react negatively to opinions advanced by generals.

The difference in "assimilative projection on peers" between the Ss susceptible to peer influence and those susceptible to the generals is in the predicted direction, although it fails to reach significance. The high "mean assimilative projection on peers" in the "little change to peers" condition was not predicted.

Table 3 presents the mean scores for similarity to the authority figures in the eight groups. Analysis shows no significant effects, and Hypothesis III regarding similarity to authority is therefore not supported. Perhaps "authority" was considered too broadly to be psychologically meaningful. Reaction to authority may be different according to the kind of authority; for instance, conveyers of society's values, charismatic leaders, or parental figures.

## SUMMARY

Students in two undergraduate psychology courses were administered the same attitude scales on two occasions. The second administration was accompanied by information concerning the expressed opinions of either generals or peers. As hypothesized, those students showing little change in their attitude scale scores were found to be high on a measure of cognitive simplicity. Those students who were influenced by peers perceived themselves as more similar to peers than students who were negative to peer influence. An hypothesis relating assimilative projection on authority figures to influence by authority was not supported by the data.

## REFERENCES

1. BERKOWITZ, L., & LUNDY, R. M. Personality characteristics related to susceptibility to influence by peers or authority figures. *J. Pers.*, 1957, **25**, 306-316.
2. BIERI, J. Cognitive complexity-simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 263-268.
3. FIEDLER, F. E. The psychological-distance dimension in interpersonal relations. *J. Pers.*, 1953, **22**, 142-150.
4. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.
5. JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504-518.
6. KELLY, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
7. LEWIS, HELEN BLOCK. An experiment on the operation of prestige suggestion. In G. E. Swansen, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Henry Holt, 1952. Pp. 18-29.
8. LUNDY, R. M. Assimilative projection and accuracy of prediction in interpersonal perceptions. *J. abnorm. soc. Psychol.*, 1956, **52**, 33-38.
9. SNEDECOR, G. W. *Statistical methods*. (4th ed.) Ames, Iowa: Collegiate Press, 1946.

Received August 1, 1956.



# MAKE-A-PICTURE STORY (MAPS) TEST FINDINGS FOR RUBELLA DEAF CHILDREN<sup>1</sup>

D. MARJORIE BINDON

Canterbury University College, New Zealand

**A**LTHOUGH the Psychological Abstracts (down to 1955) do not report any personality studies done specifically on rubella deaf children (children born deaf through the mother having German measles during pregnancy), many people closely connected with the deaf are of the opinion that these children behave differently from other defective hearing children. The writer, as a trained teacher of the deaf, has both taught and lived with hearing-impaired children and her experience has led her to query this belief. It is her view that rubella deaf children and nonrubella deaf children (those born deaf through causes other than maternal rubella or who had become deaf before the age of two) seemingly display similar personality characteristics but that, as a group, defective hearing children who have not naturally acquired speech and language do differ from the hearing children.

To test this hypothesis the writer (2, 3) studied the personality characteristics of rubella deaf children by means of the Rorschach test using the Munroe checklist system of scoring (7). Her findings confirmed her hypothesis and revealed a common rigid personality pattern differentiating both the rubella deaf and the nonrubella deaf from the hearing group at well beyond the .001 level of confidence.

*This article reports findings with the MAPS test<sup>2</sup>, also used in the same investigation. The MAPS test can be adapted to avoid a frequent limitation of personality studies of deaf*

children that do not take fully into account the considerable language retardation of the deaf. When the examiner is a trained worker with hearing impaired children, and when the procedure is altered to eliminate the necessity for verbal response and the data are limited strictly to the scoring of the formal signs, the MAPS technique can seemingly be made entirely objective, quantitative, and if necessary, nonverbal in performance.

Perhaps the most serious handicap of deafness lies in the fact that it prevents the establishment of normal interpersonal relationships (cf. 1, 4, 8, 9). The fantasy productions of deaf children may also be expected to reflect this deficit. The general hypothesis as previously stated, therefore, leads to the following specific prediction to be tested here: MAPS records of rubella deaf and nonrubella deaf do not differ in the number of "normal" signs and "schizophrenic" signs; but the rubella deaf score fewer "normal" signs and more "schizophrenic" signs than their hearing counterparts.

## METHOD

### Subjects

*Rubella deaf group.* This group was comprised of 36 fifteen-year old rubella deaf children (16 boys, 20 girls) who were born between December 1, 1938, and July 31, 1939. When tested, they were attending coeducational government schools for the deaf in New Zealand which use entirely oral methods and admit both day and residential pupils. All the qualified teachers in these schools are trained at the main school for the deaf and follow similar educational procedures. The children studied<sup>3</sup> represent a rubella deaf population that lies within the extremes of intellectual functioning, since both the ineducable, or those with serious multiple handicaps, and the most intellectually able were not, at the age of 15, attending a school for the deaf.

To permit investigating whether sex, degree of deafness, or the residential factor influence test performance differentially, this group was chosen to include as many nonresidential pupils as possible, and an equal number of "partially deaf" and "totally deaf"

<sup>1</sup>Based on part of thesis in partial fulfillment of requirements for the degree of Master of Arts in Psychology, University of New Zealand.

<sup>2</sup>MAPS test (10) is a variation of the TAT, but is less structured as figures are separate from the backgrounds. S selects and places figures on each background picture in turn and then proceeds to tell his story for that particular card. Test material comprises 22 background pictures 8½" X 11", from which E normally selects 11, and a set of 67 figures identifiable by a code symbol from a figure identification card. Data derived from figure location sheet can be subjected to objective quantitative treatment whereas S's stories can be analysed in the manner of TAT protocols.

<sup>3</sup>The names of the children whose deafness was due to maternal rubella, or to other known causes before the age of two, were compiled from the original data of D. R. Hay (5).

children who could be matched for age, sex, and educational classification (above average, average, below average). "Partially deaf" and "totally deaf" refer here, respectively, to those children with average hearing losses  $>60$  decibels but  $<101$  decibels, and  $>100$  decibels in the better ear, as tested by a pure-tone audiometer in the frequencies 512, 1024, 2048. The partially deaf children could use their residual hearing to varying extents with the help of individual hearing aids. Hence, they acquired language at a faster pace and, in general, were taught separately from the severely hearing impaired of similar age and intellectual ability.

*Nonrubella deaf group.* This group was comprised of 15 fifteen-year-old children (7 boys, 8 girls) who were either born deaf through causes other than maternal rubella or who had become deaf before they were two, and were all attending the government schools for the deaf in New Zealand. The writer is not prepared to say that these children are truly representative of a nonrubella deaf population; they were the only non-rubella deaf children available whose etiology had been clearly determined.<sup>4</sup>

*Nondeaf group.* This group was comprised of thirty 15-year-old children (13 boys, 17 girls) selected from children who volunteered to be tested and were attending a coeducational, multicourse, public, nonresidential high school that was situated in the same city as the main school for the deaf. The school was selected to represent a hearing city school population that fell within the extremes of the very highest and lowest intellectual levels and whose members were reasonably well distributed throughout the various socioeconomic categories as gauged by parental occupation. A public day school was chosen rather than a private residential school or an orphanage, since the latter types of institution not only do not sample a cross section of the school population, but are also far more organized, formal, and restrictive in their supervision of the out-of-school activities of their charges than are the government residential schools for the deaf, which are run on very progressive lines.

Because of the limited number of deaf subjects available for testing, the method of matched pairs was used to investigate intergroup and deaf subgroup differences. Therefore, from the rubella deaf group of 36 children, the 30 who were most closely matched for age, sex, educational classification, and approximate socioeconomic status with a member from the volunteer normal hearing control group were selected for comparisons with hearing children.<sup>5</sup> Similarly, but with the additional criterion of degree of hearing loss, 15 children from the rubella deaf group were paired with children in the nonrubella deaf group. Further, to investigate the possible influence of the residential factor on test performance, each of the nonresidential deaf pupils (7 rubella deaf, 5 nonrubella deaf) was also matched for

sex, age, degree of deafness, educational classification and approximate socioeconomic status with an appropriate member from the residential deaf group.

In conjunction with the Rorschach test and the MAPS test, the Wechsler-Bellevue Intelligence Scale Performance Form 11 was administered to all subjects (Ss) in this study. Statistical analysis of results by means of the *t* technique for matched pairs revealed no significant differences between the compared groups or deaf subgroups in either mean IQ or mean subtest scores, the pattern of which was essentially similar ( $p > .1$  in all cases). It should be noted, however, that these mean scores for the rubella deaf group, as for the nonrubella deaf group, were all consistently lower than for the nondeaf group; that these findings were not significant may have been due to the small size of the samples. (Mean IQ of rubella deaf group was 91.6, and *SD* 12.2, as against mean IQ 101.7, and *SD* 13.6 for the nondeaf group.)

### Administration

To overcome any bias the MAPS test may have against deaf children who are retarded in language development, the experimenter (*E*) altered the test procedure so that Ss could respond nonverbally to each stimulus situation. In the modification, Ss are not asked to relate a complete story about the figures they have selected to people each background picture. Content analysis in the manner of TAT interpretation is therefore precluded, and the scoring and analysis of the test findings is limited to the formal psychosocial sign categories that Schneidman (11) found to differentiate the fantasy productions of normals from psychotics<sup>6</sup>. The test was given individually to each S in a room occupied only by him and *E*. It came as the last of the three tests required of him. The specific procedure used for all Ss, nondeaf as well as deaf, was as follows:

The *E* emptied the box containing the cutout figures onto the table at which S was seated. The S rearranged the figures face upwards. *E* sat slightly behind and to the left of S. From his particular position not only could *E* record responses without distracting S, but also a deaf S could speech read *E* with relative ease.

The background pictures: Livingroom, Street Scene, Medical, Bathroom, Dream, Bridge, Bedroom, Blank, Doorway, Cemetery, and Forest, were presented in that order. *E* issued the following instructions orally!

Have a good look at all the little pictures on the table. There is no hurry. Tell me when you are finished. . . . Now here is a big picture (first background picture placed on table in front of S). I want you to make me an interesting story by choosing pictures from any of these here (indicating

<sup>4</sup>Ibid.

<sup>5</sup>The composition of the nondeaf group in respect to age, sex, educational classification, and approximate socioeconomic status was directly comparable to that of the rubella deaf group. The writer had intended originally to match each rubella deaf child for the above mentioned variables with a nondeaf child, but she was forced to select her normal control group from volunteers of whom only 30 fulfilled her matching criteria.

<sup>6</sup>Shneidman (11) established a set of 59 objective signs (39 "normal" and 20 "schizophrenic") that could statistically differentiate these two groups. Shneidman defines a MAPS test sign as "a denotable aspect of the subject's performance, specifically having to do with his selection and handling of the test material." (11, p. 189). It refers to a structural aspect of the performance as against the content of performance and relates to what is done with the figures rather than what is said about them.



figures on table) and putting them on this picture (pointing to background picture). You can use as many of these pictures (pointing to figures) as you like. Off you go.

If *S* put just one figure on the background picture he was told, for the first card only, that he could use more than one figure if he wished. This instruction was given in case a deaf *S* had misunderstood the opening test instructions. *E* then told *S* "Yes, that's right. Now tell me about the story." (*E* was interested only in gaining information to help her to record the theme and decide whether figure interaction had taken place and was not concerned with securing content material for a possible TAT-like analysis: verbal responses were not essential for securing this information which could as readily have been obtained pantomime fashion.)

Well, that was a good story. I have some more pictures here (indicating pile of background pictures) for you to tell me a story about. You can use the same little pictures here (pointing to figures on first background picture) again, or you can use any of the others (indicating remaining figures with a sweep of the hand).

The *S* then returned the figures he used to the table, the first background picture was withdrawn and another put in its place. All the background pictures were dealt with in the same manner. Schneiderman was followed explicitly in the formal scoring of the figure location sheets.

## RESULTS AND DISCUSSION

### Quantitative Findings

To avoid capitalizing on chance differences (cf. 6, p. 245), it was decided not to compare groups on their total individual sign scores unless their composite sign score differences yielded a significant  $\chi^2 p < .01$  ( $\chi^2$  corrected for continuity,  $df = 1$ , cutting point lower median score).

Both MAPS composite sign scores, "normal" and "schizophrenic," failed to differentiate the nonrubella deaf group from its matched rubella deaf group; they also failed to differentiate the matched deaf subgroups formed on the basis of degree of deafness, sex, and residential factor, respectively. (The differences in all cases were not large enough to warrant statistical analysis.) These results are consistent with the first half of the hypothesis under investigation; namely, that rubella deaf children do not differ significantly from nonrubella deaf children in number of "normal" signs and "schizophrenic" signs on the MAPS test. Moreover, neither sex, degree of deafness, nor the residential factor appeared to have had any differential influence on the fantasy productions of the deaf population tested as measured by these composite sign scores.

On the other hand, in confirmation of the latter half of the specific prediction, the MAPS test results show that the rubella deaf sample gave significantly fewer "normal" signs ( $p < .001$ ) and significantly more "schizophrenic" signs ( $p < .01$ ) than did its matched hearing group (Table 1). Because these over-all results were significant, the scores for these groups were examined for differences in the separate sign categories. Out of a total of 59 sign comparisons, 6 "normal" signs and 3 "schizophrenic" signs were found to differentiate the above groups ( $p < .05$ —Table 1).

The deaf (rubella and nonrubella) and non-deaf *Ss* were then compared on these particular 9 signs and on their composite sign scores, the last child in the rubella deaf sample of 36 being arbitrarily dropped so as to facilitate statistical calculations. This step was taken in order to test the general hypothesis that, as a group, defective hearing children who have not naturally acquired speech and language behave differently from the hearing. The statistical findings confirmed this hypothesis. The sign category scores on which the deaf and hearing subjects were compared were, without exception, found to differentiate them at beyond the .05 level of confidence; both composite sign scores and all individual sign categories, except for *FI* (figure interaction) with spouse and the presence of *M-8* (priest) in any background (both dropped from  $p < .02$  to  $p < .05$ ) had chi squares that were larger than those calculated for the nondeaf and the paired rubella deaf group (Table 1).

### Qualitative Findings

Only qualitative information that supplements the quantitatively established differences will be discussed here. Since the qualitative performances characteristic of the rubella deaf are also characteristic of the nonrubella deaf, the discussion in this section refers to the deaf as a group.

The outstanding fact to emerge from comparison of deaf and nondeaf on the "normal" sign scores was that the group difference in each figure interaction category was highest when the interaction represented the most personal and frequent interrelationships of normal everyday life, and was lowest for the least personal and least frequent. Thus, *FI* with child by adult gave  $p. < .001$ ; *FI* with

TABLE 1  
COMPARISON OF MATCHED RUBELLA DEAF AND NONDEAF, AND OF DEAF AND NONDEAF GROUPS WITH RESPECT TO CERTAIN MAPS SIGNS\*

Category	Rubella Deaf ( <i>N</i> = 30)	Non- deaf ( <i>N</i> = 30)	$\chi^2$	<i>p</i>	Deaf ( <i>N</i> = 50)	Non- deaf ( <i>N</i> = 30)	$\chi^2$	<i>p</i>
"Normal" Signs								
"Normal" composite sign score	12	28	16.86	< .001	22	28	17.42	< .001
Child figure three or more times	2	9	4.01	< .05	5	9	14.9	< .001
Figure interaction with acquaintance	19	27	4.57	< .05	31	27	6.04	< .02
Figure interaction with spouse	12	22	5.50	< .02	22	22	5.39	< .05
Figure interaction with parent	5	23	17.87	< .001	7	23	28.8	< .001
Figure interaction with child by adult	3	21	20.07	< .001	6	21	25.75	< .001
In Medical, man takes an exam	1	11	8.44	< .01	1	11	15.06	< .001
"Schizophrenic" Signs								
"Schizophrenic" composite sign score	22	11	6.73	< .01	38	11	9.98	< .01
Feet off ground	17	5	8.68	< .01	32	5	15.48	< .001
In any background—priest, <i>M-8</i>	19	9	5.42	< .02	28	9	4.11	< .05
Ignoring background	18	1	19.72	< .001	30	1	23.04	< .001

\* In testing "normal" and "schizophrenic" composite sign scores for significant differences, lower median score for matched rubella and nondeaf groups was used as cutting point (9 and 1, respectively).

parent  $p < .001$ , *FI* with acquaintance  $p < .02$ , *FI* with spouse  $p < .05$ ; in comparison with the remaining *FI* categories where no statistically significant differences existed: *FI* with stranger and *FI* with animal gave  $p > .1$  in both cases. The lack of themes involving figure interaction on the part of the hearing-impaired was not so much due to their giving more stories to backgrounds that contained only one figure, but to their tendency to populate the backgrounds with figures each going his own way or, apparently, representing separate stories. For instance, a street scene of six figures with no interaction between any of them may be given: or, in the dream, four figures may be placed, each of whom represents a different dream that the sleeper had had.

Group differences in the selection of the same child figure three or more times ( $p < .001$ ) may indicate that the defective hearing group show less constancy in their identification with a child figure. This indication of a lack of constancy in identification seems to be supported by the investigator's observation that the individual deaf *Ss*, more than the hearing, tended to use a greater variety of figures to represent the same type of character (e.g., a father or mother); in doing this they tended, more than did the nondeaf, to include in their selection of human figures, figures outside the male adult, female adult, and child categories (e.g., minority category: Negroes,

Jews, etc., silhouette and blank faces category; and undetermined as to sex category).

Although the nondeaf sample gave a higher proportion of figure relationships involving verbal interaction, the nature of the interrelationships was, in general, very similar for both groups: Physical aggression (mainly shooting or fighting) predominated, and the most common feelings expressed were anger, fear, and to a lesser extent, sorrow, while the emotions of love, joy, or happiness were conspicuously absent.

It has already been mentioned that an outstanding characteristic of the fantasy productions of the deaf is their lack of themes involving interpersonal relationships; another outstanding feature is their element of incongruity due, apparently, to a lack of reality testing (cf. "schizophrenic" signs: ignoring background,  $p < .001$ ; feet off ground,  $p < .001$ ). Information from the figure location sheets supplied other types of inappropriateness given by the deaf but not by the hearing group. For example, specific figures were selected for specific backgrounds for which they are not normally selected: *F-1* [nude female figure] in the Doorway, *L-4* [ghost] in bath in the Bathroom; incongruous themes were given: "This man [*M-6*—man with a gun] shot that man [*M-7*—supine blood-stained male] and policeman [*M-5*—policeman standing feet astride, hands on hips] saw"; the



placing of *M-7* on the operating table in the Medical and a male figure beside it, with the theme being that *M-7* was dead and the other figure was the doctor. It is of interest to note that these two themes, or very similar variations of them, were given by several deaf children as against none of the hearing, and it is tentatively suggested that they exemplify not only incongruity, but also the assigning of ineffectual roles to people in authority.

The more frequent use that the deaf made of the schizophrenic sign *M-8* (priest, used almost exclusively in the Cemetery background) can probably be accounted for in terms of an association of ideas and does not in their case indicate a heightened religiosity (cf. Shneidman, 11).

#### SUMMARY

With procedure altered slightly to eliminate the deaf subjects' language handicap, the MAPS test was given to three groups of 15-year-old children: 36 rubella deaf, 15 nonrubella deaf, and 30 hearing children. For purposes of intergroup matched pair comparisons, members of the nonrubella deaf group and nondeaf group, respectively, were matched with individuals selected from the rubella deaf group; the groups to be compared being then (a) rubella and nonrubella deaf (15 pairs) and (b) rubella deaf and nondeaf (30 pairs).

Results (quantitative and qualitative) for both deaf groups were similar and when matched deaf subgroups were compared, neither degree of deafness, sex, nor institutional residence was found to have had any specific influence on their performances ( $p > .1$  in all cases). However, the rubella deaf, in comparison with the nondeaf, scored fewer "normal" signs ( $p < .001$ ) and more "schizophrenic"

signs ( $p < .01$ ). Six "normal" signs and three "schizophrenic" signs were found to differentiate these two groups ( $p < .05$ ).

Furthermore, a comparison of the deaf as a whole (rubella and nonrubella,  $N = 50$ ) with the nondeaf revealed that the above signs could also distinguish these two populations ( $p < .05$  in all instances).

It is contended that the fantasy productions of the deaf, in general, are indicative of their social isolation and illogical unrealistic thinking.

#### REFERENCES

1. BEST, H. *Deafness and the deaf in the United States*. New York: Macmillan, 1943.
2. BINDON, D. MARJORIE. Personality characteristics of Rubella deaf children: Implications for teaching of the deaf in general. *Amer. Ann. Deaf.*, 1957, 102, 264-270.
3. BINDON, D. MARJORIE. Rubella deaf children: Rorschach study employing Munroe Inspection Technique. *Br. J. Psychol.*, in press.
4. GETZ, S. *Environment and the deaf child*. California: California School for Deaf, 1955.
5. HAY, D. R. Relation of maternal rubella to congenital deafness. *N. Z. Med. Journal*, 1948, 48, 604.
6. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
7. MUNROE, RUTH L. The inspection technique for the Rorschach protocol. In L. E. Abt, & L. Bellak (Eds.), *Projective psychology*. New York: A. A. Knopf, 1950. Part 11.
8. MYKLEBUST, H. Clinical diagnosis of children with impaired hearing. *Volta Review*, 1946, 48, 738-742.
9. ADVISORY COUNCIL ON EDUCATION IN SCOTLAND. A report on pupils who are defective in hearing. Edinburgh: His Majesty's Stationery Office, 1950.
10. SHNEIDMAN, E. S. A manual for the MAPS test. *Proj. Tech. Monogr.*, 1951, 1, No. 2.
11. SHNEIDMAN, E. S. Schizophrenia and the MAPS test. *Genetic Psychol. Monogr.*, 1948, 38, 145-233.

Received August 7, 1956.

# SHIFT IN JUDGMENT OF WEIGHTS AS A FUNCTION OF ANCHORING STIMULI AND INSTRUCTIONS IN EARLY SCHIZOPHRENICS AND NORMALS<sup>1</sup>

KURT SALZINGER<sup>2</sup>

*Columbia University*

THE literature concerned with investigation of the variables which are effective in changing an absolute scale of judgment is a rich one (1, 3, 5, 6, 7, 9, 10, 13, 15, 16) dealing with many varied types of stimuli which impinge upon the different sense receptors of the organism. The consistency of findings has led investigators to postulate a single judgment process to explain the formation of and shift in a scale of absolute judgment with normal subjects (Ss).

In this study, differences in the judgment mechanism between normal Ss and schizophrenics were studied by using a modified form of the method of absolute judgment. The modification was employed because a pretest of the method indicated that in the presence of the anchor stimuli some Ss instructed themselves to shift their judgments and some to keep them constant. The instruction variable was therefore controlled by telling Ss not to shift. This change in the administration of the method of absolute judgment made it possible to describe Ss' judgments in the anchor condition in terms of two conflicting discriminative stimuli. The first of these, a physical stimulus, was the anchor in conjunction with the test stimulus following it. The second, a verbal stimulus, consisted of the instructions of experimenter (E).

In terms of Goldstein's conceptualization of schizophrenic thinking (2), reaction to the verbal stimulus may be viewed as an instance of abstract thought, whereas reaction to the physical stimulus characterizes concrete

thought. Amount of shift (i.e., the change in judgment from the unanchored to the anchor condition) can then be said to vary as a function of the extent to which a given S reacts to concrete as contrasted to abstract stimuli.

The shift phenomenon may be looked upon from another point of view, that of constancy. If the instruction variable is effective, then it serves to keep the judgment constant. Whether or not an anchor stimulus precedes it, the same judgment is evoked for a given stimulus. The extent to which S shifts then serves as an index of his inability to preserve constancy. In many instances, constancy is maintained when the response to abstract aspects of the stimulus or stimulus situation is prepotent over the response to the concrete aspects of the stimulus or stimulus situation.

On the basis of this formulation, schizophrenics and normals were compared on their ability to respond to concrete versus abstract stimuli, or on their ability to maintain constancy. The central hypothesis was that schizophrenics are less able to maintain constancy or to react to abstract (verbal) stimuli than normals.

Other studies have been made using similar techniques on patient groups. Sonder (14) found judgment of lifted weights to be prognostic in a study of schizophrenics. Kaplan and Helson (4) demonstrated differences in performance between lobotomized patients and normals in a pilot study on a weight judgment task. Zubin (18) used the same weight judgment technique as that employed in this study for a comparison of patients who leave the hospital with those who do not. He found that patients who stayed in the hospital shifted more than those who left.

Rado's (11) theory that proprioceptive disorder is an innate trait of schizotypal organization gave rise to the second hypothesis of this study; namely, that schizophrenics make more incorrect judgments than normals in a task where judgments depend upon proprioceptive discrimination.

<sup>1</sup>This investigation was supported in part by research grant M586C from the National Institute of Mental Health. The article is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree in the Department of Psychology in the Faculty of Pure Science, Columbia University in 1954. The author is indebted to Professors J. Zubin, H. Garrett, C. G. Mueller, Dr. S. Kugelmass, and Mrs. S. Salzinger for their invaluable aid during the various phases of the dissertation on which this paper is based.

<sup>2</sup>Now at Biometrics Research, N. Y. State Dept. of Mental Hygiene, Psychiatric Institute, 722 West 168 Street, New York 32, N. Y.



Mueller and Schumann (8) indicated that the lighter a weight, the higher and faster it is lifted, and that judgment of heaviness depends upon *S*'s response to the ease with which he overcomes each weight's inertia. The design of this experiment allowed a retest of this contention as the third hypothesis.

Both judgment time and height of lift of each weight were further explored as relevant parameters of the judgment situation. These variables were examined in order to elucidate any differences in shift found between normals and patients.

## METHOD

### Apparatus

The weight judgment apparatus consisted of a turntable which could be revolved in such a way as to make available one stimulus at a time to *S*, an arm rest allowing *S* to keep his hand in a comfortable and steady position, and seven weights of 100, 200, 250, 300, 350, 400, and 900 gms. Each of the weights was encircled with lines marking one-cm. distances from the bottom so that there was a total of seven such lines. A buzzer-chronoscope circuit was so set up that the buzzer served as a signal for *S* to lift the weight, and the time from the signal to the time of judgment could be obtained as an index of judgment time.

### Subjects

Sixteen acute schizophrenic patients (eight males and eight females) were randomly chosen from one male and one female ward of the New York State Psychiatric Institute. All patients were cooperative and in good contact. Sixteen normals (eight males and eight females), most of whom were chosen from the School of General Studies of Columbia University, were equated with patients (matched by pairs) as far as was practically feasible in terms of sex and age. The mean age of the patient group was 27.6 years (range from 19 to 38 years), while that of the normal group was 27.2 years (range from 19 to 39 years). Patients ranged in educational status from one term in high school to the Ph.D., while the education of the normals ranged from one year of college to the Ph.D. The diagnoses of the patients were: 3 schizophrenic undifferentiated, 6 schizophrenic paranoid, 2 schizophrenic mixed, 3 schizophrenic catatonic, 1 schizophrenic simplex, and 1 schizophrenic pseudoneurotic. Patients participated in the experiment as part of their regular routine while in residence in the hospital. Normals were paid for their participation in the experiment. Unfortunately, two normal *Ss* were presented with the experimental conditions in incorrect order and had, therefore, to be discarded from the sample in the second session. When normal patient comparisons were made, the patients matched with the discarded normals were left out of the analysis.

### Experimental Procedures

Both normals and schizophrenics underwent the same experimental procedures. The *E* ascertained from

each *S* whether he engaged in an occupation involving weight estimation (no such *S* was found) and whether he was right or left handed (*S* was subsequently requested to use his dominant hand). All *Ss* were presented with the following instructions upon entering the experimental room:

This is a weight judgment task. I am going to ask you to judge the heaviness of five weights in comparison to each other by calling the heaviest "very heavy," the next heaviest "heavy," the next one "medium,"—that's the middle weight—the next one "light," and the lightest one "very light."

Now suppose you place your arm in this arm rest (right for right-handed, left for left-handed). Right underneath your hand you will find the knob of a weight. When I give you the signal like this (sound buzzer), you will lift the weight approximately one inch from the table top, set it down, and give me your judgment of it. Now just to show you the correct judgments, I'll present the weights to you in order, telling you which judgment applies to each of the weights. This is the very heavy weight (sound buzzer), this is the heavy one (buzzer), this is the medium one (buzzer), this is the light one (buzzer), and this is the very light one (buzzer).

This procedure familiarized *S* with the proper categories and trained *S* to respond to the buzzer signal and to lift the weight the required distance by means of his thumb and two neighboring fingers. After this initial training period, *E* made no additional comments concerning *S*'s height of lift. At the end of the training period *E* continued:

O.K.? Now you have lifted all the five weights. From now on I'll present the weights to you in a mixed up order and you'll give me your judgment of the heaviness of the weight in terms of very heavy, heavy, medium, light, and very light. O.K.? I'll sound the buzzer, you lift the weight, set it down and give me your judgment of it.

The *E* then presented *S* with the five stimulus weights (200, 250, 300, 350, and 400 gm.) five times each in the pattern of a Latin square. Each weight was thus presented in a different position in each series of five weights. This unanchored condition was presented to each *S* at the beginning of the experiment in exactly the same way. Four male and four female normals and their matched patient *Ss* were then presented with a heavy anchor; the rest of the *Ss* were presented with a light anchor. The instructions for this anchor condition were:

From now on I'm going to ask you to lift an extremely heavy (light) weight before lifting each of the weights that you are to judge. Don't bother to judge the first weight and make sure that its heaviness does not affect your judgment of the second weight. O.K.?

In this condition, the weights were presented to *S* in exactly the same order as in the unanchored condition, except that *S* was required to lift the anchor before each weight that he judged.

At the end of the anchored session, *S* was allowed to remove his arm from the arm rest, and *E* answered questions (without disclosing the purpose of the experiment) for a period of approximately five minutes. When the rest period ended, *E* trained *S* as in the beginning of the experiment, repeated the instructions,

and then presented *S* with the unanchored condition. At the end of this period, those *Ss* who had been presented with the heavy anchor before the rest period were presented with the light anchor, while those who had been presented with the light anchor before the rest period were presented with the heavy anchor. This second anchored condition was presented to *S* with the same instructions as before.

Fourteen patients and 11 normals were retested after a lapse of two days; the other 2 patients were retested after 4 days, while 2 of the normals were retested after 3 days, 1 after 1 day, 1 after 4 days, and 1 after 8 days. All *Ss* were retested with the same procedure except that the order of the anchor conditions was reversed. The *Ss* who had previously been given the heavy anchor first were now given it second, and vice versa.

The test weights in the unanchored condition and the anchor weights in the anchored condition were presented to each *S* approximately eight seconds after the preceding judgment; the test weights in the anchored condition were presented approximately four seconds after *S* set the anchor weight down.

During the eight-second period, *E* recorded the judgment of *S*, the height of lift correct to the nearest centimeter, and the decision time correct to the nearest .01 of a second.

The decision time as measured included the time between the sounding of the buzzer and *S*'s lifting of the weight, *E*'s reaction time to *S*'s vocal judgment (disconnecting the chronoscope circuit), and *S*'s decision time proper (the time from lifting the weight to the judgment). Since the reaction times involved were much shorter than the whole decision time, however, the latter could still be used for comparative purposes as it was in this experiment.

None of the *Ss* was aware of the fact that *E* was recording either their decision time or their heights of lift.

The raw data were converted to stimulus shift scores and kinaesthetic scores as described in detail by Salzinger (12). These scores were based on the difference in grams between the weight to which a particular judgment was assigned and the one to which it should have been assigned according to the verbal instructions given during the preliminary training period.

## RESULTS

### Shift

Combining the shift scores of both test and retest sessions, one shift score was obtained for the heavy anchor and one for the light anchor for each *S*. A comparison of normals and schizophrenics using Wilcoxon's paired replicates test (17) showed that the latter shifted significantly more than the former ( $.02 < p < .05$ ) in the heavy anchor condition but did not differ from the normals in the light anchor condition ( $p > .05$ ).

For the first session alone, the results were the same. The schizophrenics shifted more than the normals ( $.01 < p < .02$ ) from the unanchored to the heavy anchor condition,

but no difference was found from the unanchored to the light anchor condition ( $p > .05$ ). For the second session alone, there were no significant differences between normals and patients for either anchor.

Since there was a significant difference in shift between normals and patients in the first session but not in the second, the changes from the first to the second session were examined. Wilcoxon's paired replicates test demonstrated a significant ( $p = .05$ ) increase in shift due to the heavy anchor for normals from the first to the second session. There was a nonsignificant trend in the opposite direction for the patients. No significant differences in shift were found for the light anchor conditions. The contrary changes in both groups were apparently responsible for the elimination of the significant difference between them.

Change in relative status of each *S* from the first to the second session was examined in terms of rank order correlation coefficients between shift scores. Statistically significant correlations ( $p < .05$  by a one tail test) were found for the shift due to the heavy anchor of both normals ( $N = 14$ ,  $\rho = +.49$ ) and patients ( $N = 16$ ,  $\rho = +.44$ ). On the other hand, correlations for the light anchor condition for both normals ( $\rho = +.11$ ) and patients ( $\rho = +.33$ ) showed no such relationship.

### Kinaesthesia

There was no statistically significant difference ( $p > .05$ ) by Wilcoxon's (17) paired replicates test between normals and patients in their kinaesthetic ability. This was tested by comparing the kinaesthetic score (total error score in grams) of the patients and their controls on all unanchored conditions. Comparison of the change in discriminative ability from the unanchored to the anchored conditions in normals with that of the patients also revealed no statistically significant difference.

### Decision Time

Decision time data were tabulated separately for each session, for each weight, and for each condition (unanchored, heavy anchor, and light anchor). Median values were then computed separately for the normals and patient. Results are shown in Table 1.

Wilcoxon's groups-of-replicates test indicated that the decision time of the patients



TABLE 1  
MEDIAN DECISION TIME (IN SECONDS) OF JUDGMENTS OF NORMALS AND PATIENTS FOR FIVE DIFFERENT WEIGHTS UNDER CONDITIONS OF NO ANCHOR (NA), HEAVY ANCHOR (HA), AND LIGHT ANCHOR (LA) IN TWO SESSIONS

Sessions	Weights	Normals				Patients			
		NA	LA	NA	HA	NA	LA	NA	HA
1	200	3.90	4.44	3.86	4.13	4.06	4.56	4.36	3.83
	250	4.43	4.24	4.05	4.26	5.17	4.67	4.42	4.09
	300	4.42	4.32	3.88	4.30	5.36	4.81	4.88	4.79
	350	3.88	4.41	4.01	4.08	4.86	4.26	4.71	4.46
	400	3.94	3.76	3.84	4.19	4.82	4.13	4.46	4.35
2*	200	3.99	4.28	3.83	4.18	3.70	3.83	3.56	3.59
	250	4.61	4.37	4.51	4.59	3.71	3.86	3.89	4.01
	300	4.73	4.69	4.51	3.86	4.15	4.03	4.41	4.25
	350	4.38	4.46	4.47	5.01	3.86	3.98	4.34	4.18
	400	4.07	4.20	5.02	4.98	3.73	3.51	3.76	4.19

\* There were 16 normals in the first session, 14 in the second session, and 16 patients in both sessions. Each S was presented each weight five times in each condition.

was significantly longer ( $p < .01$ ) than that of the normals in the first session and significantly shorter ( $p < .01$ ) than that of the normals in the second session.

Significant interaction, determined by a chi-square analogue of an analysis of variance in Wilcoxon (17), was found between groups (patient *vs.* normal) and experimental conditions (anchored *vs.* unanchored) in the first session only, i.e., the largest differences in decision time appeared in the unanchored condition. The group-condition interaction in the second session and the group-weight interaction in both the first and second sessions were nonsignificant.

Analysis of the change in decision time from the first to the second session showed a significant ( $p < .01$ ) increase in decision time for normals and a significant ( $p < .01$ ) decrease for patients by Wilcoxon's paired replicates test.

Session-condition interaction was significant ( $p < .05$ ) for the patients, the heavy anchor condition showing the smallest change from the first to the second session; session-condition interaction for normals and session-weight interaction for both normals and patients were not significant.

Differences between conditions were tested for by a chi-square analogue of analysis of variance. A significant difference was found for patients between conditions in the first ses-

sion ( $p < .05$ ) but not in the second. No significant differences were found for either of the two sessions of normals. In the first session, patients showed the longest decision times in the two unanchored conditions, the shortest in the heavy anchor condition, and the next to the shortest in the light anchor condition.

Differences between weights were tested by the same technique and yielded no significant differences for either session of normals. For patients, significant differences were found between the decision times for different weights ( $p < .05$  for first and second sessions). The longest decision time was associated with the middlemost weight in the series, the 300 gram weight, with increasingly shorter decision times for weights farther from the middlemost.

### Height of Lift

The accuracy of *E*'s estimate of height of lift was determined by having a person lift a weight to a predetermined height by means of a pulley system and then recording *E*'s estimate of height. Ten per cent of the heights were incorrectly judged by *E*. No error exceeded .5 cm.

Height of lift data were tabulated separately for each session, for each weight, and for each condition (unanchored, heavy anchor, and light anchor). Median values were then computed separately for normals and patients. The results are presented in Table 2.

The height of lift data were subjected to the

TABLE 2  
MEDIAN HEIGHTS OF LIFT (IN CM.) OF WEIGHTS OF NORMALS AND PATIENTS FOR FIVE DIFFERENT WEIGHTS AND UNDER CONDITIONS OF NO ANCHOR (NA), HEAVY ANCHOR (HA), AND LIGHT ANCHOR (LA) IN TWO SESSIONS

Sessions	Weights	Normals				Patients			
		NA	HA	NA	LA	NA	HA	NA	LA
1	200	3.3	2.0	2.9	3.5	4.3	3.5	4.0	4.0
	250	3.0	1.7	2.9	3.2	4.4	3.4	4.4	4.2
	300	2.9	1.8	2.5	3.2	4.3	3.5	4.2	4.2
	350	3.2	2.0	2.7	3.6	4.5	3.8	4.4	4.2
	400	2.8	2.1	3.2	3.0	4.2	3.6	4.4	4.1
2*	200	3.1	2.3	2.7	3.5	4.6	3.4	4.0	4.6
	250	2.6	2.3	2.5	3.2	4.4	3.6	4.3	4.5
	300	2.7	2.4	2.4	3.3	4.3	3.9	4.4	4.3
	350	3.3	2.6	2.6	3.4	4.2	3.7	4.4	4.4
	400	3.2	2.4	2.5	3.7	4.4	3.6	4.2	4.4

\* There were 16 normals in the first session, 14 in the second, and 16 patients each in both the first and second sessions.

same kind of analysis as the decision time data. Wilcoxon's groups-of-replicates test indicated that the patients lifted their weights significantly ( $p < .01$ ) higher than the normals in both sessions.

Group-condition interaction was significant at the .005 level for the first session and at the .03 level for the second session. In both sessions, the light anchor condition consistently produced the smallest differences between patients and normals. Group-weight interaction was not significant for either session ( $p > .05$ ).

Height of lift did not change significantly ( $p > .05$ ) from the first to the second session, nor was there a significant session-weight or session-condition interaction.

Analysis of differences between conditions and between weights by a chi-square analogue of analysis of variance (17), done separately for each session of normals and patients, yielded significant ( $p < .05$ ) differences between conditions but not between weights. The heavy anchor condition consistently resulted in the lowest height of lift for both normals and patients. Among normals and in the second session of the patients, the light anchor condition resulted in the greatest height of lift, the unanchored condition falling in the middle between the light and heavy anchor condition. In the first session, the patients exhibited the second greatest height of lift in the light anchor condition and the greatest height of lift in the unanchored condition.

The relationship between the degree of shift and the change in decision time from the unanchored to the anchored conditions separately for normals and patients in each session and for each anchor was explored as a factor relevant to the judgment situation. Examination of the relationship yielded no significant rank-order correlations.

#### DISCUSSION

The central hypothesis that schizophrenics shift more than normals was confirmed for the heavy anchor condition. The ineffectiveness of the light anchor compared to the heavy anchor may perhaps be accounted for by its being only 100 gms. below the lightest test weight, whereas the heavy anchor was 500 gms. above the heaviest test weight.

The instability of the differential effect of

the heavy anchor on normals and patients from test to retest required a more exact examination.

Analysis of the change in shift due to the heavy anchor showed that the lack of difference between normals and patients in the second session was primarily due to a change in the normals. They shifted significantly more in the second than in the first session, thus narrowing the initial large difference in shift between the two groups.

The correlations between test and retest showed that the shifts due to the light anchor were not reliable, while the shifts due to the heavy anchor for both patients and normals showed reasonable reliability ( $p < .05$ ).

The results of the analysis of shift may be further clarified by exploring the relationship of decision time to shift. The change in decision time from the first to the second session paralleled that of shift; there was an increase in decision time for normals and a decrease for patients. Both these changes were significant. Further examination made it clear that differences between normals and patients were largest in the unanchored conditions. From these results it would appear that resistance to the anchor may be sustained only at the expense of a longer decision time than is otherwise necessary. This interpretation is consistent with the fact that in the first session, where the patients shifted more, their decision times were shorter for the anchored than for the unanchored conditions, while in the second session, where no significant differences in decision time occurred between the anchored and unanchored conditions, the patients shifted less. This notion applies to the first session for the normals, where there was no significant difference in decision time between anchored and unanchored conditions with relatively great resistance to shift, but does not seem applicable to the second session for normals, where slight resistance to shift was accompanied by nonsignificant differences between anchored and unanchored conditions. Furthermore, correlations of amount of shift with decision time within the group of normals and within the group of patients were not significant.

The height of lift data did not shed any light on the change in amount of shift from the first to the second session since neither



normals nor patients showed differential changes from one session to the other.

It now remains necessary to explain the instability in the normal-patient difference in ability to shift. Since only the normals changed significantly in their shift scores, it may be that their motivation decreased from the first session to the second. While they followed instructions closely in the first session because of their initial curiosity, they later became bored and had difficulty resisting the anchor; consequently, they took longer for their decisions. On the other hand, patients who showed decreased decision time, but no significant change in shift, may be supposed to have been as highly motivated to do well in both sessions but more anxious to get through with the task because of its potential threat (to them). This was evidenced by the fact that the patients showed concern during both testing sessions over whether their performance would have an effect on their remaining in the hospital in spite of all reassurances to the contrary by E.

Since the correlations between the shift scores of the two sessions were reliable though low the above interpretation is offered as a tentative explanation for the lack of difference in shift between normals and patients in the second session.

A comment might be made here concerning the low correlations. In general, the psychophysical techniques are used only after intensive training of highly motivated Ss who have achieved high reliability in their judgments. When applying psychophysical techniques in an experiment involving patients, however, it becomes necessary to keep the task sufficiently short to hold the patients' attention. Repetition of the task on successive days, which in normals results in highly stable performance (learning has reached an asymptote), may encounter the typical fluctuations in the pathology of patients. It may be of interest to examine this problem by subjecting both patients and normals to longer training sessions.

The second hypothesis tested in this experiment, that patients have poorer kinaesthetic sense than normals, was not confirmed. Thus, while patients may still show poorer proprioceptive sense than normals in areas other

than the kinaesthetic one, this experiment would seem to call such a contention into question.

As far as the Mueller-Schumann theory was concerned, the evidence was inconsistent with the notion that height of lift is a cue for judgment. Neither the normals nor the patients lifted the weights to differential heights, but they were still able to discriminate the stimuli.

Exploration of other factors relevant to the judgment situation revealed a consistent difference between normals and patients in the relative amount of time devoted to the judgment of each weight. The normals appeared to use the same amount of time for each weight, while the patients took the longest time for the middlemost weight with decreasing decision times for heavier and lighter weights.

#### SUMMARY

A weight judgment task was administered to 16 schizophrenic patients and 16 normals for 2 sessions. The two groups did not differ from each other in ability to discriminate between the weight stimuli, but the schizophrenics shifted significantly more than the normals in the heavy anchor condition, thus providing evidence for the theory that schizophrenics are more prone to react to concrete than to abstract stimuli, as well as for the theory that schizophrenics are less able to maintain constancy in weight judgment than normals. Measures of decision time and height of lift were also obtained and used to throw light upon the differences in shift.

#### REFERENCES

1. BEEBE-CENTER, J. G. The law of affective equilibrium. *Amer. J. Psychol.*, 1929, 41, 54-69.
2. GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Los Angeles: Univer. of Calif. Press, 1944.
3. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297-313.
4. HELSON, H. Perception and personality. *Air Univer. USAF Sch. of Aviation Med.*, 1953. (Rep. No. 1, Project No. 21-0202-0007.)
5. HILL, C. W. Perceptual judgment as a function of mental set, anchoring point, and method of judgment. *J. exp. Psychol.*, 1953, 46, 325-328.
6. HUNT, W. A. Anchoring effects in judgment. *Amer. J. Psychol.*, 1941, 54, 395-403.

7. JOHNSON, D. M. Learning function for a change in the scale of judgment. *J. exp. Psychol.*, 1949, **39**, 851-860.
8. MARTIN, LILLIE J., & MUELLER, G. E. *Zur Analyse der Unterschiedsempfindlichkeit*. Leipzig: J. A. Barth, 1899.
9. MCGARVEY, HULDA R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol.*, 1943, **40**, No. 281.
10. POSTMAN, L., & MILLER, G. A. Anchoring of temporal judgments. *Amer. J. Psychol.*, 1945, **58**, 43-53.
11. RADO, S. Dynamics and classification of disordered behavior. *Amer. J. Psychiat.*, 1953, **110**, 406-416.
12. SALZINGER, K. Techniques for computing shift in a scale of absolute judgment. *Psychol. Bull.*, 1956, **53**, 394-401.
13. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, **27**, No. 187.
14. SONDER, SYLVIA L. Prediction of improvement in psychotic patients. Unpublished doctor's dissertation, Univer. of Texas, 1954.
15. TRESSELT, MARGARET E., & BECKER, M. Scales of judgment and personality correlates. *J. gen. Psychol.*, 1950, **43**, 221-230.
16. WEVER, E. G., & ZENER, K. E. The method of absolute judgment in psychophysics. *Psychol. Rev.*, 1928, **35**, 466-493.
17. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.
18. ZUBIN, J. Psychological prognosis in early and chronic schizophrenic patients. *U S P H Progress Report M 586 C*, 1954.

Received August 9, 1956.



# AN ANALYSIS OF INTERPERSON CORRELATIONS AMONG THIRTY PSYCHOTICS<sup>1</sup>

ROBERT J. WILLIAMS<sup>2</sup> AND VINCENT S. MACH<sup>2</sup>

*Columbia University*

**I**N RECENT years there has been a growing interest in the development of prognostic criteria in mental disorders. For the most part, early attempts in this area were concerned with the identification of prognostic factors of a clinically observable nature (3, 7). Predicting the outcome of psychotic disorders from psychological test performance is, by comparison, a fairly recent development.

Windle (11), in a review of the literature on the prognostic use of psychological tests, has observed that a majority of studies on the prognostic value of various psychological tests report that good test performance is positively correlated with outcome. However, he also noted that a number of studies report an inverse relationship between test performance and outcome. The apparent contradictions were resolved when it was demonstrated that good performance on the part of patients with an acute (short duration) psychosis, and poor performance on the part of patients with a chronic (long duration) psychosis were both related to favorable outcome (12, 15).

A test of the prognostic value of psychological measures for recovery from psychosis must take account of those conditions which intervene between test observations and outcome. It is pertinent, for example, to consider the relationship between therapy and outcome. An evaluation of the contribution of therapy to outcome must be based on the comparison of spontaneous remission rates with remission rates following the various types of therapy. There is some agreement that special therapies produce remission rates no higher than good custodial care (9, 15). However, Zubin (13) has found that special therapies hasten improvement in some cases, although in the long run the remission rate may be the same as that following custodial care. Windle (11) concludes that, with respect to prognosis, individual factors are probably more important than type of therapy.

<sup>1</sup> Portions of this article were presented at the annual American Psychological Association convention, September 3, 1956.

<sup>2</sup> Now with Alfred Politz Research, Inc.

Another intervening condition which must be considered in evaluating prognostic indices is the time which is allowed to elapse between the initial psychological test observations and the follow-up observations. Windle and Hamwi (12) noted that a considerable amount of the confusion surrounding prognostic indices may be attributed to the neglect of some observers to report the time of follow-up, as well as the duration of illness at the time of prognostic testing. Very short-term follow-ups are unlikely to offer a clear indication of the ultimate status of the patient, and of course will fail to take into account the possibility of readmission.

An opportunity to relate a variety of psychological test performances to a sufficiently distant follow-up status was provided by the psychological test and follow-up data of the Columbia-Greystone study (6). This investigation was designed originally to measure the effects of selective partial ablation of the frontal cortex. In the course of the study a great variety of psychometric observations were made on a number of chronic psychotic patients. Some were operated and some were observed as controls. Tests were made both pre- and postoperatively. The patients are still under follow-up observation.

In the present study, factorial methods were applied to the interpersonal correlations among 30 chronic psychotic patients, in order (a) to determine if, among chronic psychotics, there exists a "psychometric type" with a high probability of recovery, and (b) if so, to describe this "psychometric type."

The existence of a "recovery type" has been suggested by several investigators in this area. Balinsky (1), and Lourie, Pacella, and Piotrowski (5) argue that when scores on performance tests exceed scores on verbal tests, prognosis is more favorable than when the reverse relationship exists. Pfeffer (8), Zubin and Windle (14), and Zubin, Windle, and Hamwi (15) have hypothesized that chronic patients who tend to be conceptually clear but perceptually confused have poor prognosis. These two hypotheses are not necessarily con-

tradictory and may simply be alternative expressions of the same idea. Although the number of subjects examined in the present analysis is small, and the population from which they are drawn is difficult to define, it is hoped that this study may contribute a little to the clarification of these notions.

Because the conditions under which the findings presented here could not readily be replicated, the results must be conservatively interpreted. However, an exhaustive analysis of a small number of patients may add something to knowledge in this area.

### METHOD

#### Subjects

Thirty chronic psychotic patients of the Columbia-Greystone study were used in this analysis. Table 1 presents the age, sex, and date of first hospital admission of each patient. For the purposes of the original investigation, some of the patients were given special therapy in the form of psychosurgery. The last column of Table 1 shows whether or not each patient received this treatment. The patients are identified by the same code numbers as those used to identify the patients in the original Columbia-Greystone study. In the table, the patients are divided into two groups: those who were judged "recovered" by a criterion described below, and those who were judged "nonrecovered."

A "chronic" psychosis is defined, for the purposes of the present study, as one of two years' duration or longer. All of the present subjects (Ss) had been hospitalized for at least two years at the time of testing with a mean duration of  $10\frac{3}{4}$  years.

Nearly all Ss for whom reasonably complete histories could be obtained showed evidence of prehospitalization disturbances of behavior, either in the form of poor social adjustment (extremely labile work histories, unsatisfactory marital adjustment ending in divorce or separation, or alcoholism), or in the form of abnormal episodes (reported as "convulsions," "fits," "tantrums," "depressions" or attempts at suicide). Ten of the Ss had records of hospitalization for behavior disturbances prior to their present admission. The available personal history data for each patient are reported by Mettler (6, pp. 315-389).

#### Psychometric Materials

Sixty-six measures of psychological test performance were made on each patient. These measures were made as a part of the Columbia-Greystone project during the Spring of 1947. Only the initial or preoperative measures were used in the present study. A summary list of the measures by name is presented in Table 2. Complete details of each test and scoring procedures are available in the literature (6, 12, 15).

#### Procedure

The major steps in the analysis of the data were (a) preparation of a matrix of interperson correlations for all thirty patients, (b) factorial analysis of the matrix,

TABLE 1  
DESCRIPTIVE DATA FOR 30 CHRONIC PSYCHOTIC  
PATIENTS OF THE COLUMBIA-GREYSTONE  
STUDY

Patient Code Number	Date First Hospital Admission	Age at Time of Study	Sex	Psychosurgery
<b>"Recovered" Patients</b>				
2	Nov., 1938	46	M	Operated
4	1934	55	M	Operated
5	Feb., 1945	35	M	Operated
7	Dec., 1942	40	M	Operated
11	Jan., 1943	44	M	Operated
12	Nov., 1940	57	M	Nonoperated
13	Oct., 1944	43	M	Operated
19	Mar., 1933	50	M	Operated
21	Nov., 1926	42	M	Operated
38	1925	38	F	Operated
39	Nov., 1924	44	F	Nonoperated
<b>"Nonrecovered" Patients</b>				
1	Dec., 1915	55	M	Nonoperated
3	July, 1940	61	M	Operated
6	Nov., 1940	32	M	Operated
8	Feb., 1942	29	M	Operated
9	1928	44	M	Nonoperated
14	Jan., 1933	47	M	Nonoperated
17	Nov., 1904	60	M	Nonoperated
18	July, 1943	39	M	Operated
23	June, 1942	39	M	Operated
25	Aug., 1945	58	M	Operated
26	June, 1944	35	M	Nonoperated
29	May, 1945	20	M	Nonoperated
31	June, 1941	51	F	Operated
32	June, 1934	41	F	Operated
35	Nov., 1940	59	F	Nonoperated
36	Apr., 1945	53	F	Operated
40	June, 1943	29	F	Operated
42	July, 1928	40	F	Operated
47	Jan., 1942	31	F	Operated

(c) identification of the factors in terms of test performances, and (d) determinations of the relationship of the factors to recovery and of the interrelationships of psychosurgery, age, and sex with recovery. These steps are outlined in detail below.

*Preparation of the matrix.* The distribution of raw scores for each of the 66 measures was dichotomized at the median. Each patient was classified as scoring either above or below the median on each measure. From these classifications, a matrix of coincident above-median scores for all possible pairs of patients was prepared. This matrix was converted into a correlation matrix by converting the coincidence scores and the number of above-median scores for each patient into percentages, and entering the tables of tetrachoric correlation coefficients (4) with these values. The use of the tetrachoric correlation coefficient involves several assumptions which cannot be verified in the present data. However, the tetrachoric coefficients were found to agree very closely with several other coefficients which were computed for portions of the data and closely duplicated the hierarchical order in the original



TABLE 2  
PSYCHOLOGICAL TEST MEASURES

Test Code Number	Test Measure
1	Word Association Total Reproduction Reaction Time
2	Word Association Test Reaction Time
3	Word Association Correct Recall
4	Word Association Correct Recognition
5	Word Association Inaccurate Reproduction
6	Rorschach Total Responses
7	Rorschach Whole Responses
8	Rorschach Detail Responses
9	Rorschach Form Responses
10	Rorschach Movement Responses
11	Rorschach Color Responses
12	Rorschach Reaction Time
13	Rorschach Ratio
14	Levy Movement Cards, Scale 3 (Seeing Persons)
15	Levy Movement Cards, Scale 4 (Seeing Movement)
16	Anxiety Inventory, Full Scale
17	Anxiety Inventory, Self Descriptive Scale
18	Complaint Inventory, Full Scale
19	Social Ethical Attitudes
20	Mirror Drawing Blocking
21	Mirror Drawing Frustration
22	Complex Reaction Time Test Confusion
23	Wechsler Bellevue General Information
24	Wechsler Bellevue General Comprehension
25	Wechsler Bellevue Digit Span
27	Wechsler Bellevue Similarities
28	Wechsler Bellevue Vocabulary
29	Wechsler Bellevue Picture Arrangement
30	Wechsler Bellevue Picture Completion
31	Wechsler Bellevue Block Design
32	Wechsler Bellevue Object Assembly
33	Wechsler Bellevue Digit Symbol Substitution
34	Wechsler Bellevue Full IQ
35	Porteus IQ
36	Porteus MA
37	Porteus Q Score
38	Complex Reaction Time Test, Number of Correct Solutions
39	Complex Reaction Time Test, Number of Errors
40	Albert Test, Number Correct
41	Analogies Test, Number Correct
42	Essential Differences Test, Number Correct
43	PI Sorting Test, Part 1, Number Grouped
44	PI Sorting Test, Part 2, Verbalized Groups
45	PI Sorting Test, Part 3, Isolation of Single Items
46	Weigl Test Performance
47	Capps Homograph Test, Verbal Flexibility Score
48	Benton Visual Retention Test, Number Correct
49	PI Sorting Test, Part 4, Recall
50	Memory For Objects, Recall
52	Memory For Objects, Recognition
53	Subtraction Time
54	Addition Time
55	Critical Flicker Frequency
56	Time Judgment, Total Discrepancy
57	Time Interval Test, Reproduction Discrepancy
58	Time Interval Test, Verbal Estimation Discrepancy
59	Time Interval Test, Operative Estimation Discrepancy
60	Time Interval Test, Discrepancy sans Sound
61	Time Interval Test, Discrepancy with Sound (Buzzer)
62	Mirror Drawing, Total Sides
63	Mirror Drawing, Error
64	Meaningful Paired Associates, Number of Learning Trials
65	Semi-meaningful Paired Associates, Number of Learning Trials
66	Meaningful Paired Associates, Variability in Learning
67	Semimeaningful Paired Associates, Variability in Learning
68	Verbal Directions, Rate of Learning

coincidence scores. Ease of computation therefore favored the use of the tetrachoric coefficients.<sup>3</sup>

*The factor analysis.* The matrix of interpatient correlations was factored, using Thurstone's complete centroid method with the highest correlation in each column or row entered as an estimate of the communality for each patient. Factoring was terminated when the matrices of residuals met Tucker's criterion (2, p. 298).

At this point, criterion data were obtained for all patients. The criterion is based on the hospital status of patients as of May 5, 1954. On this date, of the 30 patients considered in this study, 17 were hospitalized, 8 were discharged, 4 had died, and one patient had been deported. Of the four who died, one died while hospitalized, and three died after discharge. The patient who died in the hospital died on May 22, 1952 after 19 years of hospitalization. This patient was assigned a hospital status of "Hospitalized." Of the three patients who died after hospital discharge, one (patient 4) died on April 9, 1952 after 4½ years of nonhospitalization, another (patient 13) died January 12, 1954 after 6½ years of nonhospitalization, and the third (patient 19) died December 4, 1949 fifteen months after leaving the hospital. Each of these three patients was assigned a hospital status of "discharged." The deported patient (patient 9) was assigned a hospital status of "hospitalized" on the basis of his deportation from Greystone in July, 1951, because of "mental illness." Thus, with a criterion of hospital status as of May 5, 1954, there was a total of 11 "discharged" or "recovered" patients, and a total of 19 "hospitalized" or "nonrecovered" patients. In Table 1, patients are classified as "recovered" and "nonrecovered" by this criterion.

An orthogonal rotation of the original factors was then carried out so as to maximize the projection of the centroid of the "recovered" group on Factor I, while minimizing its projection on all other factors.

*Identification of the factors.* Patients were selected to represent the extreme positive and extreme negative ends of each factor. Care was taken to select these patients so that their average loading on factors other than the one they were selected to represent would be a minimum. Above-median performance was assigned an arbitrary value of +1, and below-median performance an arbitrary value of -1. The performance of each selected patient on each of the 66 measures was then weighted by his loading on the factor which the patient was chosen to represent, and the products were summed over patients for each measure. These summed weights were used to rank the measures as contributors to each factor, so that the factors could then be described in terms of the tests. The factor of principal interest in the present study is the one which was rotated to pass through the centroid of the "recovered" group. The other factors were described only for the purpose of clarifying the description of this "recovery factor." These additional descriptions serve the purpose of showing what the "recovery factor" is not.

*Interrrelations of criterion data.* The rotation of one of the factors to a criterion of maximizing the contrast between recovered and nonrecovered patients tends to capitalize on chance factors which may contribute to

<sup>3</sup> Since this analysis was completed, Dr. Joseph Zubin has suggested that the coincident above-median scores, corrected for chance, might have been factored directly.

TABLE 3  
INITIAL AND ROTATED FACTOR LOADINGS OF THIRTY CHRONIC PSYCHOTIC PATIENTS

	UNROTATED				ROTATED				
	I	II	III	IV	I	II	III	IV	$h^2$
Recovered									
2	-.340	-.093	.246	.731	.360	.289	.271	.658	.720
4	.508	.513	.044	.293	.683	-.308	-.043	-.213	.609
5	.581	.184	.052	.289	.469	-.486	.002	-.029	.458
7	.383	-.264	-.046	-.160	-.200	-.449	-.029	-.036	.244
11	-.046	.382	.215	.052	.312	.171	.167	-.207	.197
12	.176	-.312	.464	.520	.216	-.270	.491	.503	.614
13	.638	-.238	-.134	.169	.090	-.681	-.131	.146	.510
19	-.230	.582	.150	-.254	.197	.411	.086	-.513	.478
21	-.322	.437	.191	-.243	.084	.449	.149	-.399	.390
38	-.180	.051	-.756	.151	.016	.187	-.748	.188	.630
39	-.591	.375	-.190	-.020	.084	.682	-.209	-.101	.526
Nonrecovered									
1	-.556	-.444	.401	-.064	-.451	.376	.478	.313	.671
3	-.506	-.262	-.253	-.437	-.623	.390	-.195	-.032	.579
6	.408	-.195	-.435	-.201	-.212	-.450	-.425	-.081	.435
8	.189	.144	.414	.145	.285	-.130	.384	-.059	.249
9	-.321	-.561	-.112	.138	-.392	.115	-.026	.531	.450
14	-.302	-.402	-.241	-.263	-.554	.151	-.175	.139	.380
17	-.422	.464	.088	.355	.464	.553	.048	.063	.527
18	-.462	.044	-.222	.348	.125	.450	-.205	.354	.385
23	-.394	-.240	.038	-.134	-.350	.291	.085	.131	.232
25	.720	.433	.307	.176	.628	-.534	.218	-.321	.830
26	.407	-.158	-.452	.172	.059	-.436	-.446	.177	.424
29	.400	-.127	.211	-.382	-.222	-.419	.207	-.314	.366
31	.223	.215	-.281	-.261	.004	-.138	-.315	-.353	.243
32	-.558	-.189	-.027	-.475	-.588	.463	.022	-.114	.574
35	-.289	-.142	.041	.056	-.130	.225	.071	.190	.109
36	.309	.702	.049	-.217	.429	-.057	-.053	-.669	.638
40	.383	-.320	.197	-.370	-.355	-.468	.218	-.180	.425
42	-.047	.341	-.225	.113	.279	.158	-.264	-.096	.182
47	.654	-.162	-.267	.141	.115	-.671	-.273	.086	.545

that contrast. Therefore, no genuine test of the significance of the relationship between this factor and "recovery" is possible. However, White's rank order test (10) was applied with this serious reservation.

The possibility that the "recovery factor" might be an artifact of age or sex differences was examined by determining the relation of these variables to the factor loadings, using White's rank test for sex and the rank-difference correlation for age. (An examination of the descriptive data in Table 1 shows that the "recovery" criterion used in this study is unrelated to duration of illness, age, sex, or psychosurgery.)

#### RESULTS<sup>4</sup>

Four factors were extracted from the matrix of interpersonal correlations. The matrix of re-

<sup>4</sup>The following tables, too bulky to be reproduced here, have been deposited with the American Documentation Institute: (a) The above- or below-median classification of each patient on each of the 66 measures, (b) the matrix of coincident above-median scores, (c) the matrix of initial intercorrelations, (d) the matrix of residuals, (e) the transformation matrix for rotating the factors, and (f) a table showing the rank, for each factor, of each of the 66 measures. Order document No. 5197 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Wash-

ington 25, D. C., remitting in advance \$1.25 for microfilm and \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

siduals after the extraction of these four factors did not warrant the extraction of additional factors according to Tucker's criterion. These factors were then rotated as described above so that the first factor would pass through the centroid of the "recovered" group. The original and rotated factors are presented in Table 3, where "recovered" and "nonrecovered" patients are separately grouped. With but one exception (patient 7), all of the "recovered" patients have positive loadings on the rotated Factor I. Of the 19 "nonrecovered" patients, 10 have negative loadings on Factor I, and 9 have positive loadings. The difference between Factor I loadings of the "recovered" and "nonrecovered" patients is significant at the .05 level by White's rank-order test ( $T = 122$ ,  $n_1 = 11$ ,  $n_2 = 19$ ). (This finding must be conservatively interpreted because of the reservations noted earlier concerning the application of any test of significance to these data.)



No relationship was found between recovery-nonrecovery and psychosurgery. A chi-square test of the independence of these two principles of classification for these patients yielded a value of 2.302 with one degree of freedom,  $P$  lying between .20 and .10.

Factor I was found to be unrelated to age ( $\rho = -.11$ , not significantly different from zero). This factor was also found to be unrelated to sex by White's rank test ( $T' = 125.5$ ,  $n_1 = 9$ ,  $n_2 = 21$ , not significant). Similarly, the criterion of "recovery" is unrelated either to age (a chi-square test of the independence of the criterion and age categories of 20-34

years, 35-49 years, and 50-64 years, yielded a value of 1.176 with 2  $df$ ,  $P$  falling between .7 and .5), or sex (chi-square of .437 with 1  $df$ ,  $P$  between .7 and .5). Neither the criterion nor Factor I is related to duration of illness at time of testing. No relation was anticipated since these patients were selected as having chronic (or long duration) psychoses.

Table 4 lists the 20 measures which contributed most to a definition of each factor. The measures are listed in a decreasing order of importance (as defined above in the Procedure section) as contributors to the factor. The minus signs in front of some test code numbers

TABLE 4  
PSYCHOLOGICAL MEASURES WHICH CONTRIBUTE MOST TO A DEFINITION OF EACH FACTOR

FACTOR I		FACTOR II	
-42 Essential Differences, Correct		-5 Word Association, Incorrect	
-48 Benton Visual Retention, Correct		-53 Subtraction Time	
-31 WB Block Design		23 WB Information	
-24 WB Comprehension		28 WB Vocabulary	
54 Addition Time		47 Capps Homographs	
-30 WB Picture Completion		49 PI Sorting, Recall	
-27 WB Similarities		68 Verbal Directions, Rate of Learning	
-38 Complex Reaction Time, Correct		-19 Social-Ethical	
-34 WB Full IQ		45 PI Sorting, Single items	
37 Porteous Q score		3 Word Association, Correct	
-23 WB Information		-8 Rorschach, Detail	
22 Complex Reaction Time, Confusion		-10 Rorschach, Movement	
-29 WB Picture Arrangement		-1 Word Association, Reaction Time	
-59 Time Interval, Operative Est.		46 Weigl Performance	
-60 Time Interval, Silent		-58 Time Interval, Verbal	
-61 Time Interval, Sound		-59 Time Interval, Operative	
-25 WB Digit Span		-60 Time Interval, Silent	
-28 WB Vocabulary		-64 Meaningful Paired Associates, Trials to learn	
19 Social-Ethical		-65 Semimeaningful Paired Associates, Trials to learn	
53 Subtraction Time		27 WB Similarities	
FACTOR III		FACTOR IV	
-15 Levy Movement Cards		-21 Mirror Drawing, Frustration	
6 Rorschach, Total Responses		-67 Paired Associates, Semimeaningful Variability	
20 Mirror Drawing, Blocking		-16 Anxiety A	
23 WB Information		-17 Anxiety B	
29 WB Picture Arrangement		-22 Complex Reaction Time, Confusion	
53 Subtraction Time		-37 Porteous Q Score	
65 Semi-Meaningful Paired Associates, Trials to learn		-47 Capps Homographs	
-2 Word Association, Reaction Time		-68 Verbal Directions, Rate of Learning	
-19 Social-Ethical		29 WB Picture Arrangement	
-37 Porteous Q		40 Albert Test	
-39 Complex Reaction Time, Errors		20 Mirror Drawing, Blocking	
43 PI Sorting, Number Grouped		-50 Memory for objects, Recall	
48 Benton Visual Retention, Correct		-39 Complex Reaction Time, Errors	
57 Time Interval Reproduction		9 Rorschach, Form	
62 Mirror Drawing, Total sides		34 WB Full IQ	
-22 Complex Reaction Time, Confusion		2 Word Association, Reaction Time	
8 Rorschach Detail		4 Word Association, Correct	
63 Mirror Drawing, Error		33 WB Digit Symbol	
-44 PI Sorting, Verbalized		31 WB Block Design	
-45 PI Sorting, Single Items		-52 Memory for Objects, Recognition	

indicate that *below*-median performance on those tests contributed to a definition of the positive end of that factor. For tests without minus signs, *above*-median performance contributed to a definition of the positive end of that factor.

The relation of Factor I to the recovery criterion shows that patients who eventually "recover" almost always have high loadings on this factor, although patients with high loadings on this factor do not invariably recover. For this reason an additional comparison was carried out. The performance of the five "recovered" patients with highest loadings on Factor I (patients 2, 4, 5, 11 and 12) was compared with the performance of five "nonrecovered" patients with highest loadings on that factor (patients 8, 17, 25, 36 and 42), using a procedure similar to that which was applied in obtaining descriptions of the factors in terms of the tests, except that the performances were not weighted by the factor loadings. It was found that the recovered patients more often than the nonrecovered show superior performance on Seeing Persons and Seeing Movement in the Levy movement cards, the Albert test, Rorschach Movement, the Picture Arrangement, Object Assembly, and Digit-Symbol Substitution subtests of the Wechsler-Bellevue, and the PI Sorting test. The recovered patients show inferior performance on Word Association Reaction Time, Time Interval Reproduction, and Rorschach Reaction Time. Nonrecovered patients, on the other hand, are superior to the recovered patients on Verbal Directions, WB Vocabulary, Porteous IQ and MA, the Analogies Test, the Weigl test, the Capps Homograph test, and Verbalized Groupings on the PI Sorting test. The nonrecovered group required fewer trials to learn Semimeaningful Paired Associates, but had greater "confusion" scores on Complex Reaction Time.

#### DISCUSSION

Factor I is associated with all of the intelligence tests. In its positive or "recovery" aspect, it is associated with poor performance on these tests as well as with long reaction times and poor time judgment. Factors II, III, and IV are also associated with "intelligence" measures, and the components of intelligence involved in these factors are more easily identified than in Factor I. Therefore, a practical

approach to a description of Factor I may be first to describe what it is *not* in terms of Factors II, III, and IV.

Factor II in its positive aspects appears to represent good memory and concept formation. It is clearly a verbal factor as indicated by the high relative importance of the General Information, Vocabulary, and Similarities subtests of the Wechsler-Bellevue in its positive aspects, and Word Association Inaccurate Reproduction on its negative side. It also represents concept formation at the verbal level as indicated by the high relative importance of the PI Sorting-Recall and Isolation of Single Items, and Weigl Test Performance. Some verbal flexibility is indicated by the Capps Homograph Score. Factor II may be thus described as containing various verbal components of intelligence.

Factors III and IV both appear to involve perceptual abilities. Factor III is associated with Rorschach Total Responses and Detail, Mirror Drawing scores (scores 20, 62, and 63), Picture Arrangement subtest of the Wechsler-Bellevue, the Benton Visual Retention Test, and Seeing Movement in the Levy Movement cards. Patients who have high loadings on this factor seem capable of good performance on the PI Sorting tests, but are unable to verbalize their groupings. Factor IV is associated with the Picture Arrangement, Block Design, and Digit Symbol subtests of the Wechsler-Bellevue and Rorschach Form responses. Certain verbal tests contribute negatively to this factor, such as the Capps Homograph test and the Wechsler-Bellevue Similarities and Vocabulary. Also contributing negatively are Mirror Drawing Frustration and both Recall and Recognition Memory for Objects. Both Factors III and IV, then, seem to contrast good "perceptual" performances with poor "verbal" performances. The first seems to deal with the perception of concrete and familiar objects, whereas the latter is concerned with the perception of somewhat more abstract designs.

Factor I, then, cannot represent the verbal and perceptual components of test performance. The positive end of Factor I, which is associated with "recovery," represents low scores on all of the various tests of "intelligence." However, there is reason to believe that poor performance results from the failure of these patients to respond to the test situa-



tion. Confusion and sluggishness is indicated by the incorrect estimation of time intervals, and the various long reaction time scores. Poor test performance, probably brought about by failure to respond to the test situation, seems then to be a necessary condition for recovery from chronic psychosis.

Among the patients who had high loadings on Factor I, the "recovered" group is distinguished from the "nonrecovered" group in that the "recovered" patients seem better able to perform well on "perceptual" tasks (Levy Movement cards, Albert test, Rorschach Movement, WB Picture Arrangement, Object Assembly and Digit Symbol, and the PI Sorting test) which require some learning in the immediate situation. The nonrecovered patients are apparently better able to make use of previous learning, particularly of a verbal-conceptual nature (Verbal Directions, Vocabulary, Porteous IQ and MA, Analogies, and Homographs). These findings lend some support to the hypothesis that prognosis is favorable when "perceptual" scores exceed "conceptual" scores (8, 14, 15). However, this hypothesis concerns inpatient comparisons and the above findings are based on outpatient comparisons. Also, because they are based on comparisons among only ten patients, they must be interpreted with some caution.

Thus, the conditions favorable for a prognosis of recovery from chronic psychosis appear to include (a) generally poor performance on tests of intellectual abilities, (b) confusion and sluggishness as evidenced in long reaction times and poor time estimation, and (c) somewhat better performance in perceptual tasks requiring learning in the immediate situation over performance on verbal-conceptual tasks which depend on prior learning.

#### SUMMARY

The purpose of this study was (a) to determine if, among chronic psychotics, there exists a "psychometric type" with a higher than chance probability of recovery, and (b) if so, to describe that type.

The subjects were 30 chronic psychotics of the Columbia-Greystone study, 11 of whom were classified as "recovered" as of May 5, 1954.

A matrix of the interperson correlations

among all subjects based on above-median scores on 66 psychological test measures was factor analysed by Thurstone's complete centroid method. Four factors were extracted. These were rotated to maximize the projections of the "recovered" group on one factor while minimizing their projections on all other factors. The conclusions drawn from the analysis are as follows:

1. Among chronic psychotics there exists a "psychometric type" with a better-than-chance probability of recovery. This type is not distinguished by age, sex, or duration of illness.
2. The type is characterized by generally poor performance on intellectual tasks, long reaction times, and poor time estimation.
3. Among patients of this type, those who recover perform better than those who do not on perceptual tasks requiring new learning, whereas those who do not recover perform better than those who do on verbal-conceptual tasks depending on previous learning.

#### REFERENCES

1. BALINSKY, B. Factors in the vocational adjustment of schizophrenics after mental hospital discharge. *J. clin. Psychol.*, 1947, **3**, 341-349.
2. CATTELL, R. B. *Factor analysis; an introduction and manual for the psychologist and social scientist*. New York: Harper, 1952.
3. CHASE, L. S., & SILVERMAN, S. Prognostic criteria in schizophrenia; a critical survey of the literature. *Amer. J. Psychiat.*, 1941, **98**, 360-368.
4. CHESIRE, L. S., SAFFIR, M., & THURSTONE, L. L. *Computing diagrams for the tetrachoric correlation coefficient*. Chicago: The Univer. of Chicago Bookstore, 1933.
5. LOURIE, R. S., PACELLA, B. L., & PIOTROWSKI, Z. A. Studies on the prognosis in schizophrenic-like psychoses in children. *Amer. J. Psychiat.*, 1943, **99**, 542-551.
6. METTLER, F. A. (Ed.) *Selective partial ablation of the frontal cortex*. New York: Hoeber, 1949.
7. PASCAL, G. R., SWENSEN, C. H., FELDMAN, D. A., COLE, M. E., & BAYARD, J. Prognostic criteria in the case histories of hospitalized patients. *J. consult. Psychol.*, 1953, **17**, 163-171.
8. PFEFFER, M. Pattern of abilities in chronic schizophrenics. Unpublished doctor's dissertation, Columbia Univer., 1952.
9. SCHERER, I. W. Prognoses and psychological scores in electroconvulsive therapy, psychosurgery, and spontaneous remission. *Amer. J. Psychiat.*, 1951, **107**, 926-931.
10. WHITE, C. The use of ranks in a test of significance for comparing two treatments. *Biometrics*, 1950, **8**, 33-41.

11. WINDLE, C. Psychological tests in psychopathological prognosis. *Psychol. Bull.*, 1952, **49**, 451-482.
12. WINDLE, C., & HAMWI, V. An exploratory study of the prognostic value of the Complex Reaction Time Test in acute and chronic psychotics. *J. clin. Psychol.*, 1953, **9**, 156-161.
13. ZUBIN, J. Evaluation of therapeutic outcome in mental disorders. *J. nerv. ment. Dis.*, 1953, **117**, 95-111.
14. ZUBIN, J., & WINDLE C. Psychological prognosis of outcome in the mental disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 272-281.
15. ZUBIN, J., WINDLE, C., & HAMWI, V. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, **21**, 342-355.

Received August 13, 1956.



# FREQUENCY AND IDIOSYNCRACY OF ASSOCIATIVE RESPONSES<sup>1</sup>

GEORGE MANDLER AND ELIZABETH W. PARNES

*Harvard University*

INDIVIDUAL differences in response to environmental stimuli can be described in terms of the frequency and the normative aspects of the responses elicited by those stimuli. Thus, the response to or perception of a stimulus is determined by the particular array of responses which that stimulus elicits from the subject (*S*). The responses which thus differentiate the stimulus from other stimuli represent not only its immediate "meaning," but can be conceptualized as the result of the *S*'s previous learning experiences with such stimuli (4). It has also been argued that this acquisition of differentiating responses forms the basic paradigm for perceptual learning (5).

Empirically, it has been shown that the frequency of responses elicited by verbal stimuli is related to familiarity with the stimulus (10) and to speed of serial and paired-associate learning (2, 7, 9). Noble (8) and Mandler (4, 6) have suggested a relationship between the associative frequency of a stimulus and its meaningfulness, theoretically defined or empirically determined. Finally, it has been shown that individual rates of learning are related to the subject's frequency of emitted responses (7).

This paper reports two studies of the consistency of individual differences in frequency of response to different stimuli and the effect of stimulus variation upon response emission. Five general hypotheses are examined:

1. There are significant differences among *S*s in both frequency and idiosyncratic content of associative responses.
2. The *S*s show consistent tendencies in the number of responses evoked by a variety of different stimuli.
3. There are consistent tendencies in the content of the responses evoked by different stimuli, defined as the relative frequency of idiosyncratic or individual-specific responses.
4. Stimulus classes differ from each other in the elicitation of responses in respect to both frequency and to content.

5. A failure-stress situation produces measurable differences in individual response output and idiosyncratic content.

It is patently impossible to elicit and to quantify free association material in a controlled laboratory situation. However, this study represents an initial attempt to evoke controlled associations to a variety of different materials. The term "controlled" rather than "free" is used since the *S* is instructed to give those responses which are *directly* elicited by the stimulus. In this manner, a first approximation to individual differences in associative material may be obtained. The materials used in this preliminary study were selected to represent a diversified range of visually presented materials. Both verbal and nonverbal stimuli were included in order to ascertain any effect which might be directly related to response differences directly attributable to the verbal nature of the stimulus materials. In addition, it was decided to include some stimuli which are part of the clinical armamentarium in order to ascertain any differences between fairly simple stimulus materials and those classically believed to represent the best conditions for the elicitation of personally meaningful material.

In line with these considerations, four classes of stimulus materials were selected. The two types of verbal stimuli consisted of nonsense syllables and adjectives. It has experimentally been shown that differences in the associative frequency of nonsense syllables are empirically useful, while adjectives were selected as representing verbal materials more closely related to the *S*'s everyday learning experiences. The nonverbal stimuli consisted of line drawings and Rorschach cards. The line drawings, ranging from simple objects to nonsense figures, again tap a more familiar array of stimuli, while the Rorschach cards represent a clinically useful and relatively unstructured stimulus situation.

In order to assess further the usefulness of this technique in assessing personality variables, it was decided to introduce a simple yet widely used experimental procedure. Failure-stress was selected for this part of the study

<sup>1</sup>This study was supported by grants from the Laboratory of Social Relations, Harvard University, and from the Foundations Fund for Research in Psychiatry.

because it has been shown to be useful in the elicitation of personality differences (3).

### METHOD

The first of two studies (Study I) investigated the responses of 20 Ss to 40 different stimuli, ten in each of four stimulus classes: Nonsense Syllables, Adjectives, Drawings, and Rorschach cards. In Study II, two groups of 12 Ss each were used. The control group responded to 24 stimuli (six from each of the four stimulus classes); the experimental group responded to 12 stimuli, were then given failure-stress instructions, then responded to the remaining 12 stimuli.

### Materials

In Study I, the stimuli were selected from each of the four stimulus classes in order to obtain a maximum range of differentiation in meaningfulness and structure on the basis of available information.<sup>2</sup>

Nonsense syllables were selected from the list of 100 published by Mandler (6). This standardization provides an index of associative frequency which describes the mean number of associations elicited during a 30-sec. period. These 100 syllables were divided into ten groups of 10 syllables each on the basis of an index of ambiguity. This index is derived from the difference between the mean number of *different* responses elicited by a stimulus and the mean number of total responses elicited. Ambiguity increases as this index approaches zero. An index of zero would be obtained if no response to a particular stimulus is given by more than one S. One syllable was chosen at random from each of the ten ambiguity groups.

Nine adjectives were selected from each of nine different frequency categories from the Thorndike-Lorge word list. They ranged from more than 100 per one million words to 4 per 18 million words. The tenth adjective was selected from an unabridged dictionary as one appearing both unusual and unfamiliar to the experimenters.

All 10 drawings were selected on the basis of unpublished data collected by Baum<sup>3</sup> from the pictures used by Heidbreder in her concept formation studies (1). The range of drawings was from simple realistic objects to abstract complex figures, and selection was based on the number of different single responses elicited by the stimuli on two subsequent trials. The selected drawings ranged from 11 per cent to 86 per cent in the percentage of different responses elicited from a group of Ss, and from 11 per cent to 73 per cent in the percentage of Ss

who changed the response from trial one to trial two, thus covering a wide range of response variability.

The ten standard Rorschach cards were also used.

In Study II, only six stimuli were used in each of the four stimulus classes. The selection procedure was identical for the four classes. The stimuli within each class were ranked on the basis of the number of responses elicited in Study I, and ranks Nos. 1, 2, 5, 6, 9, and 10 were selected. In the case of tied ranks, the stimulus with the greatest number of popular responses (defined below) was chosen. The range obtained in respect to ambiguity and frequency was highly similar to that in Study I for the Syllables, Adjectives, and Drawings. This procedure resulted in the selection of Rorschach cards Nos. 3, 5, 7, 1, 6, and 8 in the order of ranks chosen.

### Subjects

The Ss in Study I were 19 male and one female undergraduate students recruited through the University employment bureau. In Study II, the Ss were 12 male and 12 female summer school students. The control and experimental groups consisted of six male and six female Ss each, assigned at random.

### Procedure

**Instructions.** The Ss in both studies were initially informed that they were to take part in an experiment designed to ascertain "what kind of responses people give to certain stimuli." They were instructed to respond to each stimulus with all the "words that the stimulus brings to mind." Ss were asked not to give chain-associations, but rather to be sure to give responses directly elicited by the stimulus. They were also requested not to give phrases or sentences, but to restrict themselves to single words.

In Study II, the control group proceeded after the initial instruction without any further interruptions. The experimental group was interrupted after the 12th stimulus presentation. At that time the experimenter (E) informed them that they were being interrupted because they were doing so poorly and that they either showed low ability or were "not trying at all." They were asked to buckle down on the rest of the stimuli. No other information was given them. They then proceeded with the second half of the stimulus presentation.

**Stimulus presentation and assignment.** The stimulus materials were presented by means of a modified juke box mechanism, which allowed the E to select the serial presentation in a predetermined order. Stimuli were mounted on 10-in. discs and presented to the S through a mirror arrangement. All stimuli were presented for 30 sec. with an interval between stimuli of approximately 10 sec.

In Study I, each S received a different order of stimulus presentations. The order of the 40 stimuli was random except that each block of four successive stimuli contained one item from each of the four stimulus classes, and no two items of the same stimulus class followed each other.

In Study II, twelve such series were prepared. Thus, the 12 control Ss and the 12 Ss in the experimental group received identical serial presentations, with the additional provision that Ss receiving the same stimulus order were of the same sex.

<sup>2</sup> The following materials have been deposited with the American Documentation Institute: Lists of stimulus materials used; the initial correlation matrix for the factor analysis; tables of response frequency and idiosyncratic percentage for the four stimulus classes for the 20 Ss in Study I. Order Document No. 5198 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting \$1.25 for 35 mm. microfilm or \$1.25 for 6 x 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

<sup>3</sup> Thanks are due Dr. Marian H. Baum, of Oberlin, Ohio, for making these data available.



### Dependent Variables

The responses of all Ss were recorded and transcribed. For purposes of analysis, a response was defined in the following way:

1. Comments to the *E*, exclamations, connectives, and other obviously irrelevant responses were excluded.

2. When the *S* responded with a phrase (contrary to the instructions), the response was categorized by the main noun, if any; otherwise, by the main verb. The main word was defined as appearing in the major clause of a sentence.

3. For counting replications of a response by *S* on a particular stimulus, a response was tallied only once if the replications represented plural or singular forms of the same noun, different forms of the same verb, or verbs and adjectives with identical stems. Thus, a single response is defined as a single word, complex word, or phrase. For coding purposes, responses were reduced to single words.

These responses were tabulated separately for each stimulus. In order to establish frequency distributions of responses given to a particular stimulus by all Ss, responses were combined, i.e., counted as replications, in a manner similar to that used for replications by Ss.

In Study I, the following measures were obtained from these tabulations for each *S*:

F—the number of responses given.

I—the number of idiosyncratic responses given, defined as those responses which are given to a particular stimulus by one, and only one, *S*.

I%—the percentage of idiosyncratic responses out of the total number of responses.

P—the number of popular responses given, defined as those responses which are given to a particular stimulus by 40 per cent or more of the Ss.

P%—the percentage of popular responses out of the total number of responses.

P'—the number of popular responses given as initial responses to the stimuli.

T—total popularity measure of responses. *T* for any one *S* is defined as the sum of the index of popularity of all responses. This index is the number of Ss giving that response. Thus, the *T* value for an idiosyncratic response is one, of a response given (to a particular stimulus) by 12 Ss is 12.

T/F—mean popularity index.

X%—percentage of responses which are neither popular nor idiosyncratic.

A—index of popularity (see *T*) of all initial responses given by *S*.

M—the number of specific responses which are given by *S* to more than one stimulus.

M'—*M* times the number of times it is given by *S*.

M'/%—the percentage of M' responses.

In Study II, only the *F* and I% measures were used.

### RESULTS

The analysis of the data proceeded in three general steps. The first consisted of a factor analysis of the response measures listed in the Method section. On the basis of this analysis,

the two variables of response frequency and percentage of idiosyncratic responses were selected for more detailed testing of the five experimental hypotheses.

### Factor Analysis<sup>4</sup>

The 13 variables listed above were used to obtain a correlation matrix based on the scores of the 20 Ss used in Study I.<sup>5</sup> For all measures, the scores computed for each *S* were obtained from all the 40 stimuli used for all Ss. Two factors were obtained, and one orthogonal rotation resulted in the factor loadings shown in Table 1.

The two factors can be relatively easily identified. Factor A shows the highest loadings for number of popular responses (*P*), number of popular responses given initially (*P'*), popularity of all initial responses (*A*), the mean index of popularity of all responses (*T/F*), and a high negative loading in the percentage of idiosyncratic responses (*I%*). This cluster of variables leads to the designation of the first factor as a popularity dimension, i.e., the tendency to give common responses or to give a low percentage of idiosyncratic responses. Factor B shows the highest loadings in the frequency of response (*F*), total number of idiosyncratic responses (*I*), and the tendency to repeat identical responses to different stimuli (*M* and *M'*). All of these variables can be conceptually related to Ss' general frequency of response output, and the factor can be designated as a frequency dimension. On the basis of these two factors, one variable was selected from each of the two for further analysis. Frequency of response (*F*) was selected as representative of Factor B and percentage of idiosyncratic response (*I%*) as representative of Factor A. The selection of *I%* was based not only on the high negative loading this variable showed on Factor A, but also because it provided a more convenient scoring system for Study II than the other popularity measures.

### Response Frequency

The individual difference variable showed a high degree of subject-to-subject variability in both studies. In Study I, the mean number of

<sup>4</sup> Dr. Ray Hyman's assistance in this analysis is gratefully acknowledged.

<sup>5</sup> See Footnote 2.

TABLE 1  
ROTATED FACTOR LOADINGS FOR ALL  
VARIABLES IN STUDY I

Variable	Factor Loadings	
	Factor A	Factor B
F	+.31	+.89
I	-.11	+.91
I%	-.75	+.44
P	+.86	+.18
P%	+.60	-.66
P'	+.86	+.20
T	+.66	+.65
T/F	+.78	-.55
X%	+.70	-.07
A	+.86	+.12
M	+.14	+.96
M'	+.13	+.97
M'%	-.05	+.69

responses elicited by a stimulus in a 30-sec. period was 6.9.<sup>a</sup> Subject means ranged from 2.6 to 16.0 with a standard deviation of 3.63. In Study II, the mean was 6.0, range 2.5 to 15.1, standard deviation 3.52. Thus, this variable provides a sensitive differentiation between individuals, with the most responsive Ss producing as much as six times as many responses on the average than the least responsive Ss.

Analyses of variance, summarized in Table 2, were performed on the frequency data from both studies. In both studies, the largest contribution to variance was from differences between Ss. Stimulus classes and stimulus differences within classes (in Study I), however, were highly significant sources of variation. Additional analysis of Study I revealed that the four stimulus classes chosen showed greater variability between classes than within classes ( $F = 6.47, p < .005$ ). Similarly, differences between Ss obviously contributed more to variance than differences between the stimuli used. Finally, the interaction between individuals and stimulus classes in Study I was significantly greater than the interaction between individuals and stimuli within classes. This finding indicates that not only do Ss react differently to different kinds of stimuli, but that this differentiation is greater than the differential reaction to stimuli within a particular class.

The absolute differences in frequency of responses elicited by different stimulus classes were not, however, very great. Table 3 shows

<sup>a</sup> See Footnote 2.

the means of the various classes, and the ranges within these classes in Study I. While the means show relatively small differences between stimulus classes, the full range for the stimuli from 4.8 (for the nonsense syllable "Duj") to 8.8 (for the adjective "Cordial" and the drawing of a house) suggests substantial differences between stimuli in their power to elicit a greater or lesser frequency of response. A surprising finding was the relatively small range of differentiation of Rorschach cards when compared with the other stimulus classes.

In Study II, where stress was introduced, all Ss' responses to stimuli within a particular class were used as the unit of response. In addition to the major effects found already in Study I, a significant source of variance was associated with the first and second halves of the experiment. This effect must be viewed in the light of the significant Halves  $\times$  Conditions interaction. It will be recalled that all Ss were presented with the same stimulus conditions controlled for serial presentation. The experimental group, however, was subjected to failure stress after the first half of the stimuli had been presented. Thus, any effect due to the introduction of this variable should show up in the Halves  $\times$  Conditions interaction. This significant source of variance can be attributed to a significantly greater increase in response frequency in the experimental group. While the control group showed a mean increase of 13 per cent, the experimental group showed an increase of 30 per cent. One of the effects of stress, therefore, was to increase the S's rate of responding well beyond a general increase in responsivity from the first to the second half of the experimental situation. The only other significant finding was the quadruple interaction between Stimulus Classes, Sex of Ss, Conditions, and Halves ( $p < .05$ ). Without prior predictions it is difficult to interpret such a complex interaction effect.

The final question to be raised concerning the frequency variable concerns the consistency of response frequency across different stimulus classes. Table 4 shows a measure, the index of concordance ( $W$ ), which indicates this consistency tendency.  $W$ 's were computed for all Ss in Study I, for the control group in Study II, and separately for the pre- and post-experimental stimuli in the experimental



TABLE 2  
SUMMARY OF ANALYSES OF VARIANCE OF RESPONSE FREQUENCY IN STUDIES I AND II

Source	Study I				Study II			
	df	Mean Square	F	p	df	Mean Square	F	p
Total	799	17.1	—	—	191	122.0	—	—
Between Ss	19	533.1	48.86	<.001	20	789.3	22.41	<.001
Stimulus Classes	3	63.2	5.79	<.005	3	300.3	18.37	<.001
Stimuli within Classes	36	9.8	3.69	<.001	—	—	—	—
Ss × Stimulus Classes	57	10.9	4.11	<.001	—	—	—	—
Halves*	—	—	—	—	—	—	—	—
Halves × Conditions†	—	—	—	—	1	740.0	21.39	<.001
					1	153.0	4.42	<.05

\* "Halves" refers to the first and second half of Study II.  
† "Conditions" refers to the experimental vs. control groups in Study II.

TABLE 3  
MEAN AND RANGE OF RESPONSE FREQUENCY (F)  
ELICITED BY FOUR STIMULUS CLASSES

Stimulus Class	Mean	Range
Nonsense syllables	6.2	4.8-7.3
Adjectives	7.1	6.1-8.8
Drawings	7.5	6.5-8.8
Rorschach cards	6.7	6.4-7.4

group. It is apparent that in all cases the index was very high, ranging from .872 to .943. This finding indicates that Ss who gave a high frequency of responses to any one stimulus class also tended to give a high frequency of responses to any other stimulus class. This tendency did not seem to be affected by the introduction of the experimental stress variable.

### Idiosyncratic Responses

It will be recalled that the percentage of S's responses which that S, and no other S, gave to a particular stimulus was chosen to represent the popularity factor derived from the factor analytic investigation. In Study I, the relationship between F and I% was found to show a nonsignificant rank-order correlation of +.19. In Study II, however, this relationship, using all 24 Ss on the first half of the stimuli, was +.72 ( $p < .01$ ). This discrepancy led to a further examination of the results of Study I. It was discovered that the low positive correlation could be attributed largely to the behavior of two discrepant Ss. These individuals showed the two lowest ranks on frequency of response and the two highest ranks on percentage of idiosyncratic responses. Using only the remaining 18 Ss, the rank-order correlation between the two measures increased

TABLE 4  
CONSISTENCY OF INDIVIDUAL DIFFERENCES ACROSS  
FOUR STIMULUS CLASSES

Source of Data	F*		I%†	
	W	p	W	p
Study I (20 Ss, 40 stimuli)	.900	<.01	.672	<.01
Study II				
Control (12 Ss, 24 stimuli)	.872	<.01	.691	<.01
Pre-experimental (12 Ss, 12 stimuli)	.943	<.01	.315	.30
Postexperimental (12 Ss, 12 stimuli)	.882	<.01	.644	<.01

\* F = Response Frequency.  
† I% = Percentage of Idiosyncratic Responses.

to +.63 ( $p < .01$ ). A further investigation of these two Ss lent support to the hypothesis that this negative correlation between frequency and percentage of idiosyncratic responses may be diagnostic of deviant adjustment. The following two quotations from the E's postexperimental notes illustrate this point:

Subject A: "... he said he hoped his answer hadn't been too 'vague' ... his vocabulary is broad but he sometimes couldn't remember what a word meant ... he wants to know more about psychology, about himself and normality ..."

Subject B: "... a bit paranoid, at least suspicious and unsure of self ... might be witty, clever in a hostile way ... probably deep-seated problems and well entrenched defenses ..."

These impressionistic comments tend to bear out the supposition of some pathological processes separating these Ss from the rest of the sample. As far as their response behavior was

TABLE 5  
PERCENTAGE IDIOSYNCRATIC RESPONSES (I%) FOR  
THE FOUR STIMULUS CLASSES

Stimulus Class	Total I%		Ss' range in I% in Study I
	Study I	Study II	
Nonsense syllables	54%	58%	22%-91%
Adjectives	53%	54%	26%-83%
Drawings	51%	55%	25%-83%
Rorschach cards	62%	57%	37%-83%

concerned, they were very inexpressive individuals, apparently tending to inhibit many of their responses. The resultant behavior was highly idiosyncratic; both Ss gave 76 per cent idiosyncratic responses. Comparable individuals either gave a large number of responses with a high percentage of idiosyncratic ones, or else gave few responses, most of which would tend to be given also by other Ss.

The range of I% showed large variability. In Study I, for all 20 Ss, I% varied from 30 per cent to 76 per cent.<sup>7</sup> The percentage of idiosyncratic responses was surprisingly stable from stimulus class to stimulus class, as shown in Table 5.

Once again there were large individual differences in this variable, and relatively small differences between stimulus classes. As in the frequency variable, it is surprising that the Rorschach stimuli did not appear qualitatively distinct from the other stimulus classes. They elicited somewhat more idiosyncratic material, but only by a few percentage points.

The effect of the experimental conditions on the relative frequency of idiosyncratic responses was similar to the effect on response frequency. An analysis of covariance of I% scores of the control and experimental groups showed a greater increase in the experimental group at the .05 level of significance. When total numbers of idiosyncratic responses were examined, it appeared that the general increase in frequency from the first to the second half differed for the two groups. In the control group, only one per cent of the response increment in the second half of the experiment consisted of idiosyncratic responses, whereas in the experimental group, 91 per cent of the additional responses were idiosyncratic. Thus, it appears that the effect of failure stress was to

<sup>7</sup> See Footnote 2.

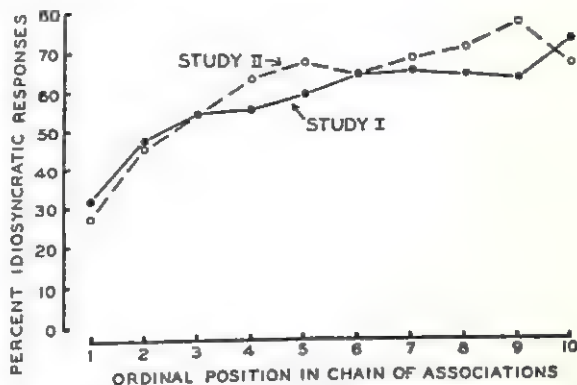


FIG. 1. PERCENTAGE OF IDIOSYNCRATIC RESPONSES ELICITED BY SUCCESSIVE ASSOCIATIONS. EACH POINT REPRESENTS THE TOTAL I% ELICITED BY ALL THE INITIAL, SECOND, THIRD, ETC., RESPONSES GIVEN BY ALL Ss

increase the productivity of idiosyncratic material.

Table 4 shows the index of concordance for I% in the various groups. Once again, though not as marked as for the frequency variable, there was a high degree of consistency from stimulus class to stimulus class. The only exception was the pre-experimental group, where consistency was not significant. These same Ss, however, seemed to achieve a significant degree of consistency following the failure stress. The nonsignificant finding prior to the stress may be a function of the relatively small number of stimuli used in determining this measure. In any case, the increase in *W* as a function of the experimental variable seems to indicate that Ss show greater consistency in their post-stress responses or, in other words, that individual differences are more emphasized as a function of the stress.

The data also provide evidence concerning the degree of idiosyncratic responding as a function of successive associations. Figure 1 shows the total percentage of idiosyncratic responses elicited by each of the first ten associations given to all stimuli. Since there were no differences between subgroups in Study II, the data from all 24 Ss were combined. The figure shows that while the initial response to the stimuli was idiosyncratic in about 30 per cent of the cases, this percentage rose steadily to about the 60 per cent level by the fifth or sixth response given. Little increase occurred after that point. These data indicate that while Ss on the whole tend to give primarily



nonidiosyncratic material during their first few associations, their later associations tend to be primarily idiosyncratic. However, no increase in the percentage of such material is obtained by the addition of more responses. This suggests that in a word association task, for example, less than 10 responses to any one stimulus are adequate to reach a high level of elicitation of personal, idiosyncratic material.

### DISCUSSION

The findings bear out the suggestion that a controlled association situation can be used profitably in describing individual differences in the frequency and idiosyncrasy of readily available associative material. The general hypotheses have been substantiated:

1. Differences between Ss show a wide range of values and constitute a highly significant and consistent source of variance.
2. Individuals show highly consistent modes of responding to a variety of different stimuli. Both the frequency of response and the idiosyncratic content of these responses show repeatable orderings of Ss across stimulus classes.
3. For the particular stimulus classes used, there are consistent differences in the amount of material these groups of stimuli elicit. While these stimuli were not specifically chosen to provide for a wide range of response variation between classes, they do represent a rather wide variety of visually presentable materials. The particular antecedent learning conditions which make for differentiations between these stimulus groups are of no particular concern to the present investigation. However, the question might be raised whether these differences are pervasive and repeatable under different conditions or whether the particular differences between classes found might be restricted to the particular situation and mode of presentation used in these studies.
4. A failure-stress situation produces measurable differences in response frequency and idiosyncratic content. It appears that the Ss in Study II responded to the report of failure and the exhortation to do better by producing more associations, and that these additional associations were primarily of an idiosyncratic character. It is, of course, open to question whether this represents a reaction to stress or anxiety per se, or whether it is a function of

the particular instructions used. The Ss' immediate reaction may well have been a conscious attempt specifically to produce more idiosyncratic responses. Taking part in a psychological experiment may well imply that one is expected to produce personal associations, while failure in a free responding situation is easily interpretable as a criticism of slow or inadequate response rate.

It has been indicated above that the Rorschach cards did not, on the basis of the measures used, represent a qualitatively different set of stimuli. Only on the percentage of idiosyncratic responses did they show a slight tendency to elicit more such materials than the other stimulus classes. This limited finding need not constitute an indictment of inkblots as useful stimulus material. The response categories utilized are extremely limited, and an extended content analysis of the material elicited by the different stimulus classes might indicate some advantage from the responses given to the Rorschach cards in terms of personality theory concepts. An impressionistic survey of the responses, however, does not substantiate this hypothesis at first glance. Rather, it appears that if the responses to the Rorschach cards tend to be useful in that sense, the S is also likely to give similar content to the other stimuli. This does suggest that diagnostic devices might well include more mundane stimulus materials than the inkblots.

The finding of deviant behavior in the cases of two Ss in Study I suggests that the present method might be refined for diagnostic purposes. Deviance in terms of idiosyncratic thought processes, such as in schizophrenia, might be predicted to show just such high idiosyncratic response frequencies coupled with low rates of responding as was found in these cases. However, such speculation must await further studies of controlled associations in particular pathological populations.

In these studies the response variables have not been related to any other personality measures. The main reason is that this investigation was considered primarily exploratory in character. Some theoretical notions both from experimental and personality studies have suggested that the associations freely available to a subject are theoretically and pragmatically useful in describing other aspects of his behavior. These studies have shown in a limited

context that these responses prove to be consistent and measurable aspects of individual differences. Rather than relate them to other personality measures, when a rationale for such relationships is lacking, it is suggested that further investigation of the antecedent conditions and behavioral correlates of these response measures would be more relevant and in the long run more useful.

### SUMMARY

Two studies were conducted to test five general hypotheses about the frequency and content of verbal associations elicited during a 30-sec. period. In the first study, 20 Ss responded to 40 stimuli, 10 each from four stimulus classes: Nonsense Syllables, Adjectives, Line Drawings, and Rorschach cards. In the second study, the effect of a failure-stress situation on response frequency and content was tested with a total of 24 Ss.

Two variables, the frequency of response and the idiosyncrasy of response (defined as the percentage of individual-specific responses) were selected for detailed study on the basis of a factor analysis of 13 response variables. The main findings were as follows:

1. Differences between Ss in frequency and idiosyncrasy showed a wide range of values. Individual differences in response frequency contributed the most significant source of variation.

2. There were highly significant consistencies in frequency and idiosyncrasy within Ss across stimulus classes.

3. There were significant differences between stimulus classes in the frequency of responses

elicited. Variability between stimulus classes was significantly greater than within stimulus classes.

4. The failure-stress situation tended to increase the productivity of idiosyncratic material.

The results indicate the utility of "controlled" associations as an individual difference variable and their possible use as a diagnostic tool.

### REFERENCES

1. HEIDBREDER, EDNA. The attainment of concepts: I. Terminology and methodology. *J. gen. Psychol.*, 1946, **35**, 173-189.
2. KIMBLE, G. A., & DUFORT, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50**, 361-368.
3. LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293-317.
4. MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235-244.
5. MANDLER, G. Perceptual learning: Stimulus differentiation of differentiating response. *Psychol. Rep.*, 1955, **1**, 79-82.
6. MANDLER, G. Associative frequency and associative prepotency as measures of response to nonsense syllables. *Amer. J. Psychol.*, 1955, **68**, 662-665.
7. MANDLER, G., & HUTTENLOCHER, JANELLEN. The relationship between associative frequency, associative ability, and paired-associate learning. *Amer. J. Psychol.*, 1956, **69**, 424-428.
8. NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421-430.
9. NOBLE, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437-446; 1952, **44**, 465.
10. NOBLE, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89-98.

Received August 27, 1956.



# THE RELATION OF EGO CONTROL TO OVERT AGGRESSION AND DEPENDENCY

NORMAN LIVSON AND PAUL H. MUSSEN<sup>1</sup>

*University of California*

**R**ECENT research indicates that aggression and dependency are, in large part, learned drives and that individual differences in strength of these drives are related to early learning experiences. Thus, the strength of the young child's dependency drive has been shown to be a consequent of amount of feeding frustration during infancy and degree of maternal nurturance and punishment during the preschool years (10), while maternal discipline and the extent of the child's identification with an aggressive like-sexed parent appear to be related to the strength of the child's aggressive drive (8, 10).

The amount of overt aggression and dependence the child manifests, however, is a function not only of the strength of his drives but also of the strength of the inhibitions against the expression of these drives. In the course of becoming socialized, i.e., acquiring the culturally customary and accepted forms of behavior, the child learns to conform with the taboos against the expression of anger, sexual impulses, and tendencies to be dependent (5). Individual differences among children in strength of inhibition of these impulses have been shown to be related to social class membership. For example, in the American middle class, strong pressures for the inhibition of aggressive and sexual behavior are applied relatively early in the child's life (6, 11). Independence training in this class is also initiated when the child is very young, although it is completed at a late age. Thus, compared with lower class children, those of the middle class are more likely to withhold overt manifestations of aggression, sexuality, and dependency. Knowledge of class differences, however, tells us little about how the child develops control of his drives. To understand this process, it is necessary to focus upon the individual experiences of the child.

According to psychoanalytic theory, the internalization of relatively strong inhibitions against the expression of these impulses may be

attributed to the functioning of the ego, the organized, executive subdivision of personality. This theory maintains that, as the child becomes aware of the reality principle and as his ego begins to develop, he acquires a rudimentary ability to test reality, to tolerate or bind tensions and to postpone gratifications. These abilities, which are major aspects or components of ego-control capacity or ego strength, presumably enable the individual to control his impulses and to cope adequately with frustrations.

With further development of ego-control capacity, there is increased recognition of, and adjustment to, culturally defined reality. Social drives become relatively more influential, compared with primary drives and immediate impulses, in guiding the child's perception of the social environment and in directing his social behavior. A child who acquires a high degree of ego-control capacity is likely to manifest more socialized behavior—behavior which conforms with socially accepted standards—than a child whose ego-control capacity is less highly developed.

Based on these theoretical considerations, the present investigation was designed to test the general hypothesis that individual differences in ego-control capacity are related to degrees of inhibition of aggression and dependency behavior.

It was first necessary to obtain measures of generalized ego-control capacity sufficiently sensitive and objective to provide reliable quantitative indices of individual differences in ego control in a group of normal children. Block and Martin (3) have presented two experimental techniques that elicit behaviors which permit assessment of two important facets of ego-control capacity. They obtained measures of nursery school children's abilities to delay gratification and to bind tension in two procedures. These measures were related to children's tendencies to regress and to attack the barrier in a replication of the Barker, Dembo, and Lewin experiment (2). In general, the results supported their hypotheses

<sup>1</sup> The authors wish to acknowledge their gratitude to Dr. Ernest Haggard for his helpful criticism of this manuscript.

that lack of ego control in children is associated both with a decrement in play constructiveness and with a tendency to attack the frustrating barrier directly.

The rationale underlying Block and Martin's use of the delay of gratification and cosatiation techniques stemmed from Lewinian conceptions, specifically those dealing with permeability of boundaries between need systems. However, it is also possible to view both tasks as frustration procedures and, hence, to consider the individual's performance as an index of his frustration tolerance—a basic aspect of his ego-control capacity. The reasoning underlying this formulation seems immediately apparent for delay of gratification; for behavior in the cosatiation task, it draws upon Burton's (4) finding that satiation may be psychologically the equivalent of frustration. Whatever the rationale for the use of these techniques, if they yield indices that provide valid measures of general internalized ego control, they should relate to the child's ability to inhibit the *overt* expression of socially disapproved drives in everyday life.

#### Hypotheses

We propose to test the following two specific hypotheses in this study:

1. Among middle class children, there is a negative relationship between children's ego-control capacities and the amount of overt aggression they express in the nursery school situation.
2. In this group, there is also a negative relationship between ego-control capacity and degree of overt dependent behavior in nursery school children.

#### METHOD

##### Subjects

Subjects (Ss) were 36 children, 19 boys and 17 girls, ranging in age from 38 to 59 months. Of this group, the three-year-olds attended the morning session, and the four-year-olds the afternoon session, of the nursery school at the Institute of Child Welfare of the University of California. The sample is a very select one: almost all of the children come from highly educated, upper-middle class families.

##### Measures of Ego-Control<sup>2</sup>

The two techniques for measuring ego control, cosatiation and delay of gratification, were adopted with slight modifications from Block and Martin (3).

<sup>2</sup> We are indebted to Dr. Frances Orr for lending the apparatus for these procedures.

The first step in the cosatiation procedure was to show the *S* how a number of spools could be packed in a rectangular rack on top of a box and then be made to fall into the box by sliding the rack. After the spools had fallen, the rack could be refilled and the same procedure repeated. The child was then told that he could play the game as long as he wished and he began operating the sliding rack himself. As soon as he indicated that he did not want to perform the task anymore, i.e., when he was satiated, the experimenter (*E*) recorded the number of trials of spool packing. *E* then placed colored blocks into the rack, telling the child he could play with these, as he had with spools, as long as he wished. The number of trials for this second performance was recorded.

The extent of deterioration in performance on the second task, following satiation on the prior one, may be expressed by an index of cosatiation very similar to the one used by Kounin (7). Performance on the first task serves as the basis of comparison: trials to satiation on the first task (spools) minus trials to satiation on the second (blocks) is divided by trials to satiation on the first task. This was the Cosatiation score. A high Cosatiation index is regarded as indicating relatively low ego control, i.e., satiation in the first task resulted in poorer performance in the second.

The experiment on delay of gratification permits direct quantitative assessment of the child's ability to delay immediate gratification of needs in order to attain greater rewards in the future. At the beginning of the experiment, *E* showed the child a small bowl containing a variety of small toys (commercially marked as bracelet "charms"). The *S* was then shown a bag full of small packages wrapped in wax paper and was told that each package contained a toy like those displayed in the bowl. It was explained that he would be given all the toys he accumulated. However, once stopped, he could not begin again.

The apparatus and procedure for acquiring the toys was then demonstrated to the child. The apparatus was a toy known as a "coke loader" which consists of a crank-operated cart that could be moved up an incline until it tipped to empty the contents. The procedure was to put one toy at a time into the cart and allow *S* to operate the crank until the toy fell into a box. After insuring that *S* understood how the toy worked, *E* again told him that he could win as many toys as he liked and, at his request, could stop playing and be given all the toys in the box. If the child had not voluntarily stopped by the time he had accumulated 20 toys, *E* ended the session by saying that time was up.

It was assumed that the child was in a conflict situation during this procedure; that he would want to claim his packages quickly to discover what they contained and to play with the toys, while at the same time, that he would want to get more toys. Presumably, the child who generally has a greater capacity to delay gratification would be able to control his desires to have his packages immediately and to examine their contents. Hence he would be expected to accumulate a relatively large number of toys before deciding to stop. The delay of gratification (Delay) score was simply the number of packages accumulated in the box.



### *Measures of Overt Aggression and Dependency<sup>3</sup>*

These were based on reports of observations of children's nursery school behavior by the three teachers in each (morning and afternoon) session. Twelve children in each session were selected for special observation for a period of two weeks or ten school days. This procedure was repeated for two more groups of 12 in a subsequent two-week session. Of the total of 48 children observed, 36 were tested in both ego-control situations, and comprise the sample of this study.

Two types of records were made: daily behavior reports for aggression and dependency, and weekly rankings of the 12 children being observed with respect to these two characteristics. The daily behavior report for aggression consisted of a checklist of 17 items of aggressive behavior. Included were: interfering with others' activities; teasing or insulting; threatening; quarrelling, screaming or swearing; physically attacking; throwing objects; destroying something intentionally; disobeying; temper tantrums. For dependency behavior there was a twelve-item checklist that included seeking physical contact, asking for help, seeking attention and/or approval, needing security object, wanting to go home, revealing hurt feelings, "helpless" crying, and asking permission unnecessarily. Each teacher filled out these two checklists for each child observed on that day.

The incidence of aggressive and dependent behavior as indicated by the total number of items checked by one or more teachers on each day was tabulated for each child over the entire observation period. Unfortunately, as a result of children's absences from school, not all Ss were observed for the same number of days; consequently, the number of daily behavior reports available was not the same for all Ss. In order to obtain comparable measures, the total incidence of aggressive and dependent behaviors for each S was divided by the number of days he was observed, yielding aggressive Acts/Day and dependent Acts/Day scores. These scores were not computed for the three Ss who were absent from school for more than five of the ten observation days. The remaining 33 Ss had aggression scores ranging from 0 (least) to 4.3 (most), with a median of .30; while the dependent Acts/Day scores ranged from 0 to 3.4, with a median of .65.

At the end of each week of observation, each of the three teachers independently ranked the 12 children she had been observing that week with respect to overall overt expression of aggression and dependency. Thus, there were six weekly rankings on each trait for each child which were summed to obtain his Pooled Ranks scores for each of the two behaviors.

The interrater reliability for the Pooled Ranks scores was determined in the following way. For each trait, each teacher-observer's average rank for the child, based on the two weekly rankings she assigned him, was treated as a score. The resulting scores of the morning and afternoon groups were then ranked separately. The rank-difference correlations between

each teacher in a session and the other two teachers in that session were computed. The median coefficient for aggression (based on six coefficients, three from each of the two sets of observers) was .80, for dependency, .48. In order to increase the low reliability of the dependence Pooled Ranks score, the analysis excluded those cases about which the observers showed the most disagreement. Specifically, if the average ranks on dependency assigned to S by any pair of observers differed by more than three units, no score was computed for him. This procedure reduced the number of Pooled Rank scores for dependence from 33 to 22.

### RESULTS AND DISCUSSION

In order to test the two major hypotheses, the measures of ego control were related to the indices of overt aggression and dependence. Since the distributions of scores on Delay and Cosatiation were non-normal and included a relatively large number of tied scores, the sample was dichotomized into High and Low ego-control groups on each of the measures, the dichotomization point in each case being selected so as to provide as nearly equal numbers as possible in the two groups. (Since High ego control is indicated by a *low* score in Cosatiation and a *high* score in Delay, the terms High and Low will henceforth be used to refer only to ego-control status.) For Cosatiation (range,  $-.20$  to  $.10$ ; median, zero) the High ego-control group ( $N = 19$ ) thus constituted included scores of zero and below; the Low group ( $N = 17$ ), scores above zero. On the Delay measure, the 20 Ss who achieved the maximum score of 20 trials comprised the High group, while the Low group ( $N = 16$ ) included the remaining Ss whose scores ranged from 5 to 18.

Composite ego-control categories were devised in order to test the combined predictive ability of the two ego-control measures which tended to be positively, though not significantly, related ( $P = .20$  by a  $U$  test comparing the actual Cosatiation scores of the High and Low Delay groups). The High Composite group ( $N = 11$ ) included those Ss who were classified as High ego control on both Cosatiation and Delay; the Low Composite group ( $N = 8$ ) was composed of Ss assigned to the Low groups on both measures.

In view of the dichotomization of the ego-control measures and the non-normality of all four of the behavior score total distributions, the  $U$  test (9) was employed to compare rank

<sup>3</sup> We wish to thank E. Peters and H. Tyau, head teachers, and E. Benjamin, E. Boydston, E. Elson, and H. Shimomaki, nursery school assistants, for their cooperation in this aspect of the study.

transformations of all overt behavior scores of the Highs and Lows on each measure of ego control. (Figure 1 summarizes all these data and shows each S's ego-control status—High or Low—as derived from the two experimental measures and his aggression and dependency scores. The legend in upper right portion of Fig. 1 indicates ego-control status on both Cosatiation and Delay measures, *not* actual scores. "H" and "L" in key refer to ego-control level on the two measures. Orientation of the Acts/Day and Pooled Ranks scales are the same; in both cases, the upper end of the scale denotes greater amount of the behavior. Reference to the appropriate portions of this figure will serve to illustrate the nature of the comparison made by a given *U* test.) The results

of these tests and their significance levels for the experimental hypotheses are summarized in Table 1.

Compared with Ss high in ego control, as measured by the Cosatiation score, those low in this measure of ego control scored significantly higher on both indices of overt aggression. There was also some tendency for those who showed relatively high ability to delay gratification to express fewer aggressive Acts/Day. For Pooled Ranks-Aggression, the difference between High and Low ego-control groups on Delay was in the same direction, though the difference was not significant. When the two ego-control measures were combined into the Composite score, the less controlled Ss were found to be significantly more aggressive than the more highly controlled Ss. Thus, there is general support of the first hypothesis, that ego-control capacity is positively related to ability to withhold expression of aggression.

Since the data from children of both sexes and of two age groups were combined in the analyses, these relationships might be attributable to some extent to covariation of the measures of ego control and aggression with sex and/or age. Thus, it is possible that girls develop greater ego-control capacity earlier and are also less likely to express aggression overtly. Similarly, both ego control and inhibition of overt aggression may increase with age. In our sample, girls did show significantly fewer aggressive Acts/Day but were not significantly different from boys on either of the ego-control measures. None of the measures of aggression or ego control showed significant relationships

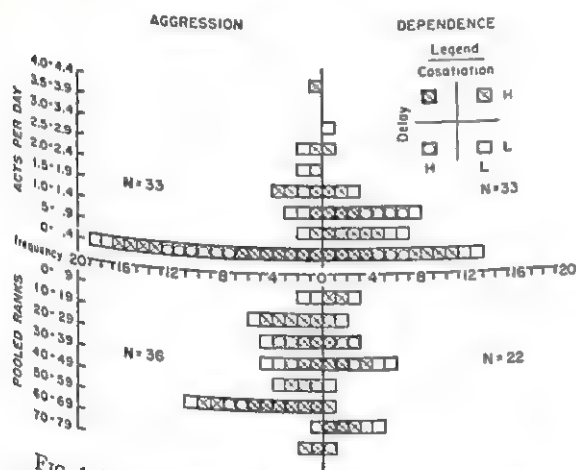


FIG. 1. DISTRIBUTIONS OF AGGRESSION AND DEPENDENCE SCORES FOR HIGH AND LOW EGO-CONTROL GROUPS. EACH UNIT IN HISTOGRAM REPRESENTS AN S. (See text for explanation of legend.)

TABLE 1  
COMPARISONS OF DISTRIBUTIONS OF AGGRESSION AND DEPENDENCE SCORES BETWEEN HIGH AND LOW EGO-CONTROL GROUPS

Behavior	Measure	Ego-Control Measure								
		Cosatiation			Delay			Composite		
		N*	U**	p***	N	U	P	N	U	P
Aggression	Acts/Day	33 (17, 16)	86.5	.04	33 (18, 15)	105.5	.14	16 (9, 7)	12.5	.02
	Pooled Ranks	36 (19, 17)	82.5	<.01	36 (20, 16)	138.5	>.20	19 (11, 8)	17.0	.01
Dependence	Acts/Day	33 (17, 16)	127.0	>.20	33 (18, 15)	111.5	.20	16 (9, 7)	24.0	>.20
	Pooled Ranks	22 (10, 12)	54.5	>.20	22 (10, 12)	58.5	>.20	12 (5, 7)	13.0	>.20

\* First number is the total *N* in the comparison; numbers in parentheses refer to the partitioning of this total into the *N*s for the High and Low ego-control groups, respectively.

\*\* All *U* values in this table are less than the value to be expected by the null hypothesis (one half the product of the two subgroups *N*s) and are in the predicted direction.

\*\*\* All *P* values are based on one-tailed tests as is required by the hypotheses.



with age. Hence, partialling out sex or age should not attenuate the reported relationships.

The data did not support the second hypothesis concerning the relationship between ego control and the overt expression of dependency needs. Although all six comparisons between High and Low groups on the three measures of ego control yielded differences in the predicted direction, none were statistically significant. This failure to detect relationships between ego-control and dependency may be due to several factors. It is quite possible, in the first place, that dependency is inherently difficult to observe and measure under conditions of free observation in a nursery school situation. The original low correlations among the observers on the dependency rankings attest to this difficulty. Another factor to be considered is that these children come from a highly select and homogeneous group of upper-middle class, highly educated families, a very large proportion of the parents having advanced academic or professional degrees. In such a group "getting along with others" is an important social value and is often considered a criterion of adequate social adjustment. Hence, socialization of aggressive impulses, especially in interpersonal situations, is likely to be emphasized from early childhood. While children in this group may be encouraged to acquire independence gradually, these parents are also likely to be permissive with respect to the expression of dependency for a relatively long period of time. Consequently, the ego-control capacity of the preschool child may have little bearing upon the amount of dependent behavior he manifests.

Our data offer some indirect evidence to support the hypothesis that there is less emphasis on the control of dependence than of aggression in this sample. Examination of the frequency distributions for the Acts/Day scores in Figure 1 shows that the incidence of both aggression and dependence generate J curves which Allport (1) has demonstrated to be typical for behavior in a number of conformity-demanding situations. He has reasoned that the steeper the slope at the high conformity end of a scale, the stronger are the conformity-producing pressures influencing the particular behavior. Dependence exhibits a

more gradual slope than does aggression, with the relative number of very high conformists being greater for aggressive Acts/Day. This difference may be quantified by a comparison of the skewness of the two distributions; the skewness indices ( $3 \text{ [Mean-Median]/Standard Deviation}$ ) are 2.78 for aggression and only 1.06 for dependence. This difference in skewness is consistent with the hypothesis of differential conformity pressures for the control of these two drives.

The data indicating that ego-control capacity is positively related to the inhibition of aggression in interpersonal relationships are consistent with the finding of Block and Martin (3) that children high in ego control make relatively few direct attacks on a frustrating barrier in an experimental situation. Thus, it is possible to interpret the findings of the present study—as Block and Martin viewed their results—as generally supportive of a Lewinian theoretical formulation.

It is also of interest to note that in both studies, performance in the cosatiation situation proved to be a better predictor than did delay of gratification scores. We can only suggest that the Cosatiation score may have been a purer, less confounded index of ego control than Delay. It measured the child's ability to continue to perform without extrinsic rewards even though he had just been satiated and was free to stop whenever he wished. Such behavior seems to correspond closely to our definition of internalized ego control. In contrast, since the delay situation provided extrinsic rewards, the final Delay score may have represented a confounding of internalized ego-control factors and a desire to obtain the available rewards. Thus, a High ego-control score on this task may have been due either to ability to delay gratification, to high motivation to possess many toys, or to both.

The results of the present study extend the meaning and implications of the concept of ego control by demonstrating how it may be useful in understanding and predicting one aspect of socialization, the inhibition of aggressive expression. Furthermore, some of the findings suggest that the concept has predictive utility only for behavior to which socialization pressure has been applied. Thus the absence of a significant relationship between ego-control measures and overt dependency in these data

may be due to the fact that there is less emphasis on dependency training in this group of Ss than we had initially assumed. For similar reasons, we would not expect to find relationship between ego control and the inhibition of overt aggression in preschool children from a social class group in which the expression of overt aggression is not discouraged at this age. Conversely, in groups which stress early training in the suppression of overt dependency behavior, our original hypothesis of negative correlation between ego-control capacity and the incidence of dependency manifestations should be confirmed.

#### SUMMARY

This study tests two hypotheses concerning the relationships between ego-control capacity in children and their ability to inhibit the overt expression of aggression and dependency needs. Thirty-six nursery school children were scored on indices of aggressive and dependent behavior manifested during a two-week observation period. These scores were related to two measures of ego control derived from performance in a delay of gratification and in a cosatiation situation.

There is a significant negative relationship between amount of expressed aggression and ego-control as assessed by the cosatiation measure. The delay measure is also negatively related to aggression, though not significantly. A composite ego-control score, involving both measures, was also related to aggression.

Dependent behavior is not predictable from any of the ego-control indices. The argument

is made that ego-control capacity should be expected to relate to the degree of expression of a drive only when that drive has been subjected to socialization pressures.

#### REFERENCES

1. ALLPORT, F. H. The J-curve hypothesis of conforming behavior. *J. soc. Psychol.*, 1934, 5, 141-183.
2. BARKER, R. G., DEMBO, TAMARA, & LEWIN, K. Frustration and regression: an experiment with young children. *Univer. Iowa Stud. Child Welf.*, 1941, 18, Whole No. 1.
3. BLOCK, JEANNE, & MARTIN, B. Prediction of behavior of children under frustration. *J. abnorm. soc. Psychol.*, 1955, 51, 281-285.
4. BURTON, A. The aggression of young children following satiation. *Amer. J. Orthopsychiat.*, 1942, 12, 262-267.
5. CHILD, I. L. Socialization. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 665-692.
6. DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child-rearing. *Amer. sociol. Rev.*, 1946, 11, 698-710.
7. KOUNIN, J. S. Experimental studies of rigidity I. *Charact. Pers.*, 1941, 9, 251-272.
8. LEVIN, H., & SEARS, R. R. Identification with parents as a determinant of doll play aggression. *Child Develpm.*, 1956, 27, 135-153.
9. MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, 18, 50-60.
10. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. psychol. Monogr.*, 1953, 47, 135-234.
11. WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale Univer. Press, 1953.

Received August 30, 1956.



# ATTITUDE CHANGE THROUGH REWARD OF VERBAL BEHAVIOR<sup>1</sup>

WILLIAM ABBOTT SCOTT

*University of Colorado*

**P**UBLIC opinion researchers are familiar, through anecdote if not personal acquaintance, with the woman who, when asked her opinion on social security, replied that she didn't know because no one had ever asked her, and she never knew what her opinions were until she opened her mouth and expressed them. Perhaps this apocryphal respondent was displaying more insight into the psychological processes underlying her attitude than is readily apparent. Though cognitive approaches to attitude development are more flattering in their emphasis on the rationally cognized relation of an object to the person's goals (1, 3, 8, 9), nonrational determiners may be equally important in the acquisition of attitudes.

According to Doob (4) an attitude may be regarded, like a habit, as an implicit anticipatory response which mediates overt behaviors, and arises out of them through response reinforcement. Thus, it is conceivable that an opinion be expressed initially, in the absence of a supporting attitude, but if the verbal behavior is rewarded, the corresponding attitude may develop and mediate subsequent opinion expressions in the presence of similar cues. This response reinforcement formulation thus focuses more on reward of verbal expressions than on cognized means-end relationships as critical factors in attitude development.

The two formulations, and the sets of variables they employ, are by no means incompatible. It is reasonable to suggest, for instance, that lacking a well structured and salient set of cognitive associations regarding an object, a person may be more readily induced to respond to it in nonlogical fashion, the outcome of his response then contributing to a nascent cognitive structure; but given a set of values and cognitions that can readily be applied to a new object, these will influence the person's attitude.

<sup>1</sup> The author is indebted to Prof. Maurice P. Smith for his suggestions regarding design and report of the study, and to the following graduate teaching fellows who assisted in the collection of data: H. Hess, P. Khanna, A. Mathews, and I. Richardson.

dinal set and thus produce a response consistent with the pre-existing cognitive structure. It is not the purpose of this paper to propose a systematically inclusive formulation of these two classes of determinants, but rather to offer evidence concerning the efficacy of response reinforcement in the determination of attitudes.

Other studies (2, 5, 10) have demonstrated the influence of social approval and disapproval on the modification of verbal behavior, but they provide no information about their effects on the underlying attitudes. The present study was designed to explore, in a controlled fashion, the effects on attitudes of reward and punishment of expressed opinions. Specifically, it was predicted that subjects (Ss) rewarded by group approval for expressing opinions opposite from their initial attitudes would show a change in the direction of the expressed opinions, while Ss punished by group disapproval for expressing contrary opinions would not show such a change. This prediction implies that reward of a new behavior increases the relative strength of the underlying predisposition, while punishment of a new behavior leads the S to revert to his formerly preferred response disposition.

## METHOD

All students in 29 general psychology discussion sections were administered free-response questions on mimeographed sheets to assess their attitudes toward three controversial issues:

1. *Universal military training.* It has been proposed that all physically able males between the ages of 18 and 25 be required to spend two years in the armed forces so that they will be trained and ready for armed service in the event of a national emergency. How do you feel about this proposal?

2. *Night hours for women students.* Senior women at the University are not required to be back in their dorms or houses by any set time at night. Some students are requesting that all women be given the same freedom of hours. How do you feel about this?

3. *De-emphasis of football.* It has been suggested that too much emphasis is placed on football at the University, in scholarships and special treatment for players. How do you feel about this?

Each of the classes was assigned, in systematic fashion, one of the three issues for debate. The instructor read the class members' opinions on the designated

issue and selected from each class two which appeared definitely favorable and two definitely unfavorable. The four students were contacted outside of class and their cooperation requested in an "experiment to see how much they could affect the opinions of class members" by debating the particular issue; but each student would take the side opposite to his own opinion. Two pairs of debators were formed for each class, members with "pro" opinions taking the "con" side of the argument, and vice versa. It was explained that, following the debate, the quality of performance would be assessed in two ways: first, by a class vote on which member of each pair did the better job; and second, by a retest of the class' opinions on the issue, "to see in which direction opinions are influenced."

The debators were counseled to avoid mentioning to anyone that they didn't really believe what they were to say. (This was intended to help prevent the occurrence of interfering responses (see 7) which might negate the effect of the experimental stimulus.) In pairing the debators, an attempt was made to equate their verbal abilities, as evidenced in their written opinions, in order that a subsequent false report of the vote would appear valid.

The debates were held two weeks after the original assessment of attitudes. One pair of opponents was called into the classroom, while the other pair waited in an adjoining room. The "pro" debator spoke first, for three minutes, followed by a three-minute presentation from the "con" debator. Then each had a two-minute rebuttal period, in reverse order. The first pair left the room, and the second pair of opponents entered for a similar performance. Then, with all four debators present, voting instructions were given to the class ("On the first ballot write which side—pro or con—did a better job on the first debate. On the second ballot write which side—pro or con—did a better job on the second debate").

The ballots were handed in, and the instructor made a pretense of counting them, while one of the students tallied votes (as read by the instructor) on the blackboard, under the name of each debator. It had been determined in advance for each pair of debators in all classes who would "win." "Winners" and "losers" were assigned alternately through the pairs of debators, so that half the time the "pros" won, and half the time the "cons" won.

The instructor then repeated the implications of the vote, mentioning the names of the "winners" and "losers," and explained that the next step in the judgment would be to see how the class's attitudes had been affected. All class members were given sheets of paper on which were printed the statement of the issue that had been debated—exactly like the first opinion assessment, except that only one issue appeared on the sheet. Debators were also asked to write their present opinions, "to see if there was any change in them."

After the opinion retest forms were collected, the nature and purpose of the experiment were explained, the pre- and postopinions of the debators were read, and an attempt made to reduce the threat of the situation for the "losers" and for those debators whose opinions had ostensibly shifted.

The pre- and postopinions of the debators were sub-

sequently typed on cards, after the deletion of certain passages that might indicate which were postopinions (such as "I still feel that . . ."). The cards were identified only by numbers, duplicated on the original opinion questionnaires, and the numbers assigned in "random" fashion, so as to prevent contamination in coding. Opinions on the cards were coded on a seven-point scale: 1—very pro; 2—pro; 3—pro, qualified; 4—neutral or balanced pro and con; 5—con, qualified; 6—con; 7—very con. Then, by reference to the original questionnaires, each card was identified as a pre- or postopinion of a "winner" or "loser."

Of the 58 pairs of debators initially designated, only 36 finally produced complete data. Twenty-one pairs failed to debate, due either to absences or to rescheduling of classes; the postopinion of one member of the final pair was not codeable. The 36 debates took place in 19 different classes, with six different instructors. From these classes a sample of 36 students whose original opinions were definitely pro or con, but who did not debate, was selected as a control group. Their pre- and postopinions were similarly transferred to cards with random identification by numbers, and coded on the same seven-point scale.

## RESULTS

It was predicted that debators who "won" would show an average change in the direction of their debate greater than the corresponding change shown by "losers" and greater than the average change of the control group. Data bearing on this prediction are presented in Table 1. Winners changed an average of 1.25 points (in the direction of debate) on the seven-point scale, while losers changed an average of 0.17 points in the opposite direction (more extreme than their original positions). The control group changed an average of 0.31

TABLE 1  
CHANGES OF WINNERS, LOSERS, AND CONTROLS  
FROM PRE- TO POSTTEST  
(Positive sign indicates mean change in direction  
opposite to Ss' original opinions)

	Winners	Losers	Controls
	(N = 36)	(N = 36)	(N = 36)
Number of changers (either direction)	24	16	18
Number of changers (toward opp. direction)	21	7	11
Mean change	+1.25	-0.17	+0.31
SD of change	1.77	1.17	1.37

Difference in mean changes:

Winners vs. losers:  $t = 3.97$ ;  $p < .001$ .

Winners vs. controls:  $t = 2.49$ ;  $p < .02$ .

Losers vs. controls:  $t = -1.58$ ;  $p > .10$ .



points toward the opinion opposite from their original one.

The difference between mean changes of winners and losers is significant beyond the .001 level of confidence, and the difference between winners and controls is significant beyond the .02 level of confidence. The difference between losers and controls is not significant.

An analysis of mean changes for the three issues separately, and for original pros versus original cons showed no significant differences among these groups, so the prediction was confirmed regardless of the issue and regardless of which extreme opinion the debator started with. When the *proportions of changers* in the three groups are compared, results are consistent with those for mean changes. More winning individuals shifted toward the opposite opinion than either losers or controls.

#### DISCUSSION

The experiment was designed to give both winners and losers equivalent contact with arguments for the opposite side, and to give both the experience of verbalizing these arguments as if they were their own. What distinguished the two groups was the brief experience of either group approval or group disapproval for their performances. These performances involved principally the oral presentation of arguments—verbal behavior to which they were unaccustomed. The vote of “win” is presumed to have reinforced the verbal behavior and with it the accompanying implicit responses—attitudes and cognitive support for them. The vote of “lose” presumably weakened whatever response tendencies had been established by the overt behavior or by cognitive contact with the opposite side, so that Ss reverted to their pre-existing attitudes.

If the opinion posttests be regarded as valid reflections of Ss' attitudes following the experimental stimulus, then one may conclude that social reward for expressing a new opinion tends to reinforce the concomitant attitudes, whereas social disapproval of the new behavior tends to lead to nonreinforcement of the accompanying attitudes. It appears that, in the present experiment, the brief experience of reward or punishment was a significant factor in accounting for attitude change, rather than

the more prolonged experience of cognitive contact with opposing arguments.

This finding is, however, subject to at least two interpretations. It may be that reward of a new behavior was the crucial factor, while cognitive contact in and of itself had no effect. Or it may be that contact with opposing arguments produced a change in attitude, which was then reinforced or extinguished depending on whether the first overt expressions were rewarded or punished. Since Janis and King (6) showed that contact with the opposite side tended, by itself, to produce some change, the latter interpretation appears more plausible. In order to help determine the relative importance of these two influences—cognitive contact and reinforcement—in this situation, it would have been necessary to have some Ss debate without a subsequent “winning” or “losing” experience. This experiment did not provide for such a condition. Regardless of whether one accepts a single-factor or a two-factor interpretation of the results, it appears that they speak clearly for the significance of reward in producing attitude change in this type of situation.

It should be noted that very few of the changes in opinion were of a spectacular sort. (Only seven of the rewarded Ss reversed the direction of their opinions; the other fourteen changers merely showed some weakening of their initial opinions, as indicated by a shift along the seven-point scale toward the neutral position.) Although the experience of success or failure was made as strong as possible under these restricted conditions, it was nevertheless a brief and perhaps not too potent one. Between the announcement of the “class vote” and the posttest of opinion no more than five minutes elapsed.

The degree to which the new opinions persisted was not assessed in this study. It is reasonable to suppose that effects of the experimental situation were transitory, since the Ss' initial attitudes presumably found considerable social support in their current friendship groups. In order to produce enduring attitude changes, it would probably be necessary to so change the Ss' social environments that the new behaviors would receive continued reinforcement. But transitory or enduring, the principles of attitude change illustrated in this experiment would appear to be applicable.

## SUMMARY

An experiment in attitude change was suggested by Doob's learning theory of attitude development. Seventy-two Ss were induced to engage in debates on three different issues, taking sides opposite to those which they had indicated as their own in an opinion pretest. Half of the Ss were rewarded, in predetermined order, by a purported vote which proclaimed them the better debators, while the other half were punished by presumably losing the debate. Posttests of Ss' opinions showed a tendency of the "winners" to change their opinions in the direction of their debates, while the "losers" did not change significantly. A control group of nondebators likewise showed no significant change in opinions.

## REFERENCES

1. CARLSON, E. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.*, 1956, 52, 256-261.
2. COHEN, B. D. *et al.* Experimental manipulation of

verbal behavior. *J. exp. Psychol.*, 1954, 47, 106-110.

3. CROCKETT, W. H. Attitude change as a function of cognitive differentiation and affect under conditions of norm-presentation with and without counter-arguments. Unpublished doctor's dissertation, Univer. of Michigan, 1952.
4. DOOB, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, 54, 135-156.
5. HILDRUM, D. C., & BROWN, R. W. Verbal reinforcement and interviewer bias. *J. abnorm. soc. Psychol.*, 1956, 53, 108-111.
6. JANIS, I. L., & KING, B. T., The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, 49, 211-218.
7. KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Rel.*, 1953, 6, 185-214.
8. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948. Chaps. 5-6.
9. ROSENBERG, M. J. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, 53, 367-372.
10. VERPLANCK, W. S. The control of the content of conversation: reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, 51, 668-676.

Received August 30, 1956.



# IDENTIFICATION AND OBSERVATIONAL LEARNING FROM FILMS<sup>1</sup>

ELEANOR E. MACCOBY AND WILLIAM CODY WILSON

*Harvard University*

**I**N the studies reported here, our interest is focused upon the process of "passive" learning. We wish to explore the conditions which govern the nature and extent of a viewer's learning the contents of a dramatic production which he views in a motion picture.

Modern psychology has assimilated thoroughly the lesson taught by Thorndike and Dewey, that we learn by *doing* rather than by having actions demonstrated to us. While there are a few reports in the literature of "observational learning" among animals (3, 4, 11, 12) all the leading learning theories emphasize that the to-be-learned response must somehow be made to occur in the learning organism, not just in a model, before its association with a given set of conditions can be strengthened. Yet it is evident that viewers of movies and television dramas do learn much of the content of dramatic programs, at least to the extent of being able to recall subsequently many details of the characters' actions, and they sometimes incorporate material thus learned into their own overt behavior. There are instances, for example, in which criminals have employed techniques for crime taken directly from movie portrayals (1).

Presumably, a viewer's learning of the actions of movie characters is enhanced if he reproduces these actions himself, at least covertly, during the screen portrayal. Our problem in predicting which viewers will learn how much of what content then becomes one of understanding the conditions that affect a viewer's tendency to match his own actions, at least covertly, to those a screen character portrays.

We assume that when a viewer becomes absorbed in a dramatic production, he *identifies* himself, at least momentarily, with one or more of the characters. By identification we mean that the viewer, in fantasy, puts himself in the place of a character and momentarily feels that what is happening to that character is happening to himself. In this process, we as-

sume that although he may reproduce very little of the gross motor behavior of his screen character (10) he does reproduce covertly<sup>2</sup> many elements of the behavior including the *emotions* he attributes to the character, so that when the character with whom he identifies himself is in a dangerous situation for example, the viewer feels fear, and when the character escapes danger, the viewer feels vicarious relief. We assume, too, that when a viewer puts himself in the place of a character, he orients toward the cues to which his character is responding. There may be misperceptions, of course, in which the viewer misjudges the nature of the cues to which the actor is responding, or misinterprets the actor's response, in which case the viewer learns something that is not an accurate representation of what the movie producer intended to portray.

While it is possible for a viewer to shift his identification from one character to another during a movie, most viewers probably make a choice of a character whom they consider to be the main protagonist of the story, and see the action mainly from his point of view. In many movies, the plot centers so fully upon one main character that all viewers may be expected to identify with him. But in other plots, there are two or more main characters sympathetically presented, and this makes it possible for viewers to differ in their choice of an "identificand." We assume that the nature of the content a viewer learns is in part determined by the choice of identificand, for the viewer presumably focuses his attention upon the stimuli that are relevant for his own character, and shares vicariously the experiences of that character rather than those of the people playing opposite him. Thus, if A assaults B on the screen, the viewers who are identified with A should feel anger and dominance, while those identified with B should feel fear, and these differential experiences at the time of viewing should be reflected in the nature of the

<sup>1</sup> This research was supported by a grant from the Laboratory of Social Relations, Harvard University.

<sup>2</sup> Covert practice has been shown to be fully as effective as overt practice in the learning of certain kinds of material from films. See Michael and Maccoby (6).

movie content that is learned. Of course, the matter of the ultimate incorporation of learned material into the viewer's own overt actions is a complex matter. This would depend not only upon the original learning of the material and perhaps its rehearsal in later spontaneous fantasy, but upon an appropriate situation arising in which the stored-up material can be evoked. Undoubtedly, much material assimilated from the mass media remains as an unused potential. A viewer may learn from a movie how to act as a groom at a wedding, but if he remains unmarried, the learning is never manifested in his behavior. Our concern in the remainder of this report is with the first step of learning from films; that is, with the subject's ability to recall film content verbally after a brief interval.

We expect, as noted, some individual differences in the choice of a movie character with whom to identify. We also assume that identification is a matter of degree, and that viewers differ in the "depth" of their identification with their chosen character at any particular moment. The deeper the identification with a character, the more should the viewer reproduce covertly, and therefore learn, the actions and feelings of that character.

In order to predict what a given viewer will learn, we must predict his choice of main protagonist, and the depth of his identification with that character. There are a number of presumably relevant factors. One undoubtedly is the consonance of the viewer's motives with the character's actions. If a screen character is doing what the viewer most wants to do, then, presumably, the viewer will be more likely to identify with him.

Another factor, we hypothesize, is that of similarity between the viewer and the screen character with respect to certain major social characteristics, such as sex, age, and race. The importance of similarity has been pointed out in connection with projective tests such as the TAT. Murray (9) has recommended that TAT pictures include a character of age and sex similar to that of the subject, to facilitate his projecting himself into a character. We believe that a similar process is involved in identifying with a movie character. While the fantasy sequence in a movie, unlike the TAT, is one over which the viewer has no control, the viewer *can* vary in the depth of his involvement

in the story—the amount of attention he pays to it—and we are assuming that his interest will be more deeply engaged if the story has a central character who is similar to the viewer in some important and discernible aspect of the life roles he occupies.

There are instances, of course, in which a viewer's similarity to a character facilitates identification but in which the viewer's motives are more consonant with the actions of a different character. For example, the character in the story who is engaged in the activities the viewer would most like to perform may be of the opposite sex or a different social status. Under these conditions, we cannot make a prediction about the outcome.

Another factor which should affect the viewer's choice of a protagonist and the depth of his identification is the extent of the viewer's anxiety about the actions which the character portrays. Our expectation that a viewer should choose a protagonist similar to himself should hold if the actions being carried out are not only attractive but socially acceptable. But if the actions are simultaneously attractive and anxiety arousing, then the viewer would probably choose an identificand *dissimilar* to himself (14). The vicarious enjoyment of tabooed activity may be somewhat less if is enacted by a dissimilar person, but the anxiety it arouses would be reduced as well, and more sharply (cf. 7). We suspect that anxiety over the expression of aggression is one of the sources of children's preference for animal heroes and space men.

In the present studies, we have focused primarily upon the factor of *similarity*. That is, we have attempted to study whether viewers do in fact identify themselves with the movie character most like themselves, given a movie in which the viewer has some choice of protagonist, and we have attempted to determine whether differential identification produces predictable differences in the nature of the movie content learned. We have also given some attention to the matter of "consonance" between the motives of the viewer and actions of particular characters. We have *not* studied the third factor—displacement of a preferred agent of action due to anxiety—but have instead attempted to minimize the problem by choosing movies in which actions of a sympathetic character that might be potentially



anxiety provoking are justified in the plot so that guilt or fear will be minimal. Thus, when aggressive actions are carried out by a hero, they are justified by the villainy of a "bad man" and occur in a context of social approval.

### STUDY I: SOCIAL CLASS

In studying the effects on identification of similarity between the viewer and a movie character, the first variable with which we worked was that of social class. The movie employed had two leading characters: a lower class boy and an upper-middle class boy. For this movie, our hypotheses were that:

1. Lower class children tend to identify with the lower class character and middle class children with the middle class character.

2. Viewers who identify with a character remember more of that character's words and actions than viewers who do not identify with that character.

3. Since both leading characters are boys, the predictions about identification apply more to boy viewers than to girl viewers.

Our hypotheses deal with the simple choice of a protagonist, not with the *degree* of involvement with that character once the choice has been made. However, consideration of degree would lead us to the parallel predictions that the more similar the viewer is to the movie character the more deeply he identifies with that character, and therefore the more his learning is specific to the content that is relevant to his chosen character.

We expected that our hypotheses might not apply uniformly throughout the movie. Specifically, the differences between the two groups of viewers (those identifying with the two different main characters) might be maximized either when both characters were on the screen at the same time, or when only one of them appeared on the screen. There were reasons, to be discussed, for either prediction. We therefore planned to use this study not only to test the specific hypotheses listed above, but to explore the group differences in selected scenes and for selected kinds of content.

### Method

*The Movie.* Two 20-minute episodes of a serial movie entitled *Junior G-Men* were used in the study; it was a story about adolescent spy chasers made about 1940. Its two main characters were an upper-middle

class boy and a lower class boy, both of whom were about 16 years of age. Harry Trent had his own car, lived in the best part of town, and could pilot an airplane and operate a radio transmitter. He was the leader of a group of boys who called themselves the Junior G-Men, and who were well dressed, polite, and had an elaborate club house. Billy Barton was the leader of a gang of "dead-end kids." They had an improvised club house in a junk yard and were ill-dressed and aggressive. Both Billy and Harry were admired fighters, but Harry won fights because he was a skillful boxer and Billy because he was a "slugger."

*Subjects.* The subjects (Ss) for this study were ten classes of seventh-grade students from public schools in the greater Boston area. The classes included 269 students—139 boys and 130 girls. Two or more classes were taken from each of four junior high schools serving communities that differed widely in socioeconomic level.

Each child in the study was asked to record his father's occupation at the end of his test blank, and the entries were coded on the Warner seven-point scale of occupational status (13). The sample included 94 children whose fathers' occupations fell at the first 3 steps of the Warner scale, 42 at step 4, and 126 at steps 5-7.

*Procedure.* The movie was shown to pairs of classes during a regular class period. The experimenter (E) told the children that they were about to see an ordinary entertainment film. The E said that he wanted to find out something about what kinds of things children notice and enjoy in movies, and that he would ask them a few questions about their reactions to the movie later on. The two reels shown were not consecutive, and while reels were being changed, the E described the intervening developments in the plot. When the two reels had been shown, the E said that he would come back in a week to ask some questions.

One week later, a multiple-choice questionnaire on the movie was administered. The items included questions on how well the S liked the movie, and several questions which were intended to reveal identification with particular characters, as well as test items on recall of the content of the movie. The Ss were told to make the best guess they could when they were not sure of an answer, and very few children omitted items in filling out the questionnaire.

### Results

The first question of interest is whether children from lower-class families more often identified with the lower class hero than did the children with middle-class backgrounds. As we defined the term, identification meant the viewer's putting himself in the place of a given character as he experienced the story. Clearly, it is difficult to obtain a direct measure of this process. There was an attempt made to get an indirect measure by asking several questions concerning the S's reactions to the main characters. We reasoned that a child would not

be likely to "put himself in the place of" a character whom he did not like and admire, since to do so would be to share disapproved characteristics. We also felt that the process of a viewer's imagining himself in a particular role is facilitated if the character playing the role behaves as the viewer himself would behave in a similar situation. We therefore asked three questions for our indirect measure of identification:

1. Which one of the two main characters did you like best?
2. Which of the two main characters would you like to be like?
3. Which of the two main characters do you feel is most like you?

The answers to these three items were combined into an identification index.<sup>3</sup> In the following discussion, a child is considered to be a "Billy" identifier if he chose Billy (the lower class boy) for two out of three of the above items. He is considered a "Harry" identifier if he chose Harry (the middle class boy) for two out of three of the items.

As may be seen from Table 1, the upper-middle class boys tended to choose Harry as an identificand and the lower SES boys tended to choose Billy. Among girls, a majority chose Harry, and this was true for girls at each social class level. Each of the items in the identification index contributed to the findings in Table 1. That is, each was positively related to social class among boys; the question "which character is most like you" contributed most, but the index correlated more highly with social class than did any individual item taken alone.

In Table 1, our measure of social class is a rating based on the occupation of the child's father. We asked each child not only about his father's occupation but about the kind of occupation the child himself would like to have when he was grown up. For girls this proved to be a difficult question, since a number wanted to be housewives, and those who did mention the kinds of jobs they would like to have tended to make choices such as secretary, nurse, teacher, or air-line stewardess which provided little range in aspired social class. For the following

<sup>3</sup> The three items had low positive intercorrelations, with coefficients of contingency ranging from .26 to .40.

TABLE 1  
RELATION OF SOCIOECONOMIC STATUS TO CHOICE  
OF MOVIE CHARACTER\*

	Socioeconomic Status (Warner Occupation Scale)	
	Upper Middle (1-3)	Lower Middle & Lower (4-7)
Boys		
Chose Billy	23%	56%
Chose Harry	75	40
Not ascertained	2	4
Total	100%	100%
Number of cases	(47)	(89)
$\chi^2 = 14.4$	$p = < .001$	
Girls		
Chose Billy	30%	33%
Chose Harry	64	61
Not ascertained	6	6
Total	100%	100%
Number of cases	(47)	(79)
$\chi^2 = .1$	$p = < .75$	

\* Not included in this table are 7 cases whose SES was not ascertained.

discussion of aspired social class, therefore, we have considered boys only.

It may be seen from Table 2 that, regardless of the social-class level of a boy's family, if he aspires to a relatively high level occupation, he tends to identify with Harry, the middle-class boy, in the movie. Boys who expect to have working-class occupations tend to identify with Billy, regardless of the occupation level of their fathers. Among boys with similar aspirations, there is only a little variation in choice of character that can be traced to current objective social-class membership. The conclusion is quite clear that a boy's choice of screen character is more closely related to the social-class level to which he aspires than to the level his family currently occupies.

*Identification and the recall of movie content.*  
Our hypothesis was that if a viewer is identified with a particular character in a movie, this fact should determine what events on the screen he perceives and remembers. Specifically, we predicted that viewers who identified with Billy should remember more of the things Billy said and did, and more of the things that happened to Billy, than should the Harry identifiers, and that conversely the Harry identifiers should remember more of the Harry-relevant material.

To test this prediction, we first designated the items in the multiple-choice test which



TABLE 2

RELATIONSHIP BETWEEN OCCUPATION TO WHICH BOY ASPIRES AND HIS CHOICE OF MOVIE CHARACTER\*

Present and Aspired Social Class	Number Choosing Billy	Number Choosing Harry
Present SES High (1-3, Warner's Occupation Scale)		
Aspired occupation high	7	30
Aspired occupation low†	3	0
Present SES Low (4-7, Warner's Occupation Scale)		
Aspired occupation high	14	23
Aspired occupation low	30	10
$\chi^2 = 10.8$ $p = <.001$		

\* Based on 117 boys for whom this information was available; 22 Ss did not provide the information.

† Too few of the high SES Ss aspired to low SES occupations to allow computation of  $\chi^2$ .

could be classed as "Billy content" or "Harry content." The following exemplifies a Harry-relevant item: "In the plans for the boys to go to the Torchies' headquarters and try to capture the gang before they escaped, what was Harry supposed to do?" This question was followed by a list of possible answers. A Billy-relevant item was as follows: "What did Billy say when he turned around and faced the man who had followed him out on the docks?" In all, there were 27 items that could be classified as "Billy content," and 20 items for "Harry content." The test had 67 items, some of which were not classified into either of these content categories. Each child's test form was scored separately for the number correct on each of these two sets of items. Then each of these scores was divided by the child's "total correct" score for the entire test in order to eliminate the effects of reading ability, IQ, and other factors that might contribute to total score but were not relevant to our study.

Table 3 shows that Billy identifiers remembered slightly more Billy-relevant material than did the Harry identifiers. The differences, though significant for boys,<sup>4</sup> are not large. The comparisons on Harry-relevant content show that both boys and girls who identified with Harry remembered slightly

<sup>4</sup> When the data are analyzed separately by each administration of the film, each administration being considered a replication, and the results combined, the Billy identifiers among the girls also remember more of the Billy content than do the Harry identifiers ( $p < .05$ ).

TABLE 3

RELATIONSHIP OF CHOICE OF MOVIE CHARACTER TO RECALL OF MOVIE CONTENT

	Category of Content			
	Billy Content	Harry Content	Billy Aggression†	Billy Non-aggression†
Boys				
Mean Score*				
Billy Identifiers	44.0	28.0	19.4	24.8
Mean Score				
Harry Identifiers	42.1	28.7	18.0	24.3
$t$	2.25	1.00	2.48	.75
$p$ †	.02	n.s.	.01	n.s.
Girls				
Mean Score				
Billy Identifiers	44.6	28.0	18.3	25.7
Mean Score				
Harry Identifiers	43.3	28.2	17.2	26.3
$t$	1.10	—	1.36	.67
$p$ †	n.s.	n.s.	n.s.	n.s.

\* Each S's score is the proportion of all items answered correctly which were classified into the content category in question. Thus, among Billy identifiers, 44% of the correctly answered questions were Billy content items.

† The "Billy content" items were subdivided into aggression and nonaggression.

‡ All the  $p$ 's are one-tailed unless otherwise stated.

more Harry-relevant content, but the differences are not significant.

A possible reason why these differences were so small is that the hypothesis tested does not take into account the attention a viewer devotes to his character's cues. We assumed that if a viewer is identified with a particular character he attends particularly to that character's actions and remarks. But it is probably also true that he attends to the stimulus properties of his character's environment. If, for example, the screen character is being approached by a dangerous animal, the viewer probably looks at the animal rather than the story hero. We could predict, then, that the viewer might notice and remember the cue properties of the hero's situation as well as (or better than) the actions and words of the hero himself. Thus, if the two main characters are on the screen simultaneously, and interacting with one another, each character's actions would be a cue for the other, and there should therefore be few differences in memory between the viewers who identify with each of the two characters.

To investigate this possibility, we subdivided

the test items according to whether the opposite character was present and interacting with the character asked about in the item. For boys, the Billy identifiers remembered significantly more of the Billy content when Harry was *not* on the screen with Billy. They remembered slightly, but not significantly, more Billy material when the two boys were present, but Billy's action did not involve talking to Harry. When Billy was talking to Harry, the Harry identifiers remembered the Billy content somewhat (not significantly) better.<sup>5</sup> For boys, then, the results fit the interpretation that when both characters are present and interacting, the nature of the remembered content is not a function of identification choice, since everything that happens is either the action of one's own character or a cue to which he responds.

We have shown so far that the different groups of viewers remembered much the same things from the movie but that the differences obtained favored the hypothesis that similarity of aspired-to class fosters identification with a given character, and that selecting one character for primary identification results in better learning of that character's actions and words. A major problem exists, however, in determining whether the differential learning is actually a result of identification. There is an alternative explanation that Billy identifiers remembered somewhat more Billy activity because Billy engaged in the kind of activities that were most interesting to the kind of boy who would choose Billy as an identificand, and that it was the interest in the activities *per se*, rather than identification with Billy, that produced the increment in learning. It is possible that Billy identifiers would have remembered more *Harry* material, if only the items about Harry had happened to concentrate upon his aggressive, motoric actions instead of his more restrained, verbal ones.

This alternative explanation is difficult to test, since we cannot guess what all the different sources of attraction for Billy-identifiers might be, nor do we have test items bearing upon them all. Partial tests were nevertheless possible with respect to aggressive versus non-

aggressive content. Starting with the possibility that lower-class nonmobile children had a higher level of aggressive motivation than the other children (see 5) we considered whether the fact that Billy's activities in the movie were much more aggressive than Harry's could account for the greater learning of Billy content by Billy identifiers. We divided the test items into those dealing respectively with aggressive and with nonaggressive activity. Too few items on Harry aggression were available to make a test for Harry content, but we could compare the recall of Billy-aggressive content with Billy-nonaggressive content. As Table 3 shows, the superiority of Billy identifiers in recalling Billy content applied to the aggressive items only. Parenthetically, it may be of interest that boys recalled the aggressive content better, and girls the nonaggressive content, both differences being significant at the .01 level. Thus we find that the Billy identifiers did not show an advantage in recall of all kinds of Billy content, but only those selected aspects that were presumably consonant with their own motives. This finding would not be predicted from our view of identification, for when a viewer puts himself in a character's place, he should experience vicariously *everything* that character does and feels, regardless of whether each item of the character's behavior touches upon one of the viewer's major motives.

From our existing data we cannot tell whether identification has contributed to our results, or whether we are simply observing the effects of selective perception based upon need. If both identification and need-consonance were required for maximal recall of movie material, we should find that the Billy identifiers did not recall Harry aggression especially well, even though they are presumably set to select aggression, because the aggression is not carried out by their own preferred protagonist. In the absence of test data on Harry aggression, we cannot answer the question. More evidence bearing upon this problem is available from Study II.

## STUDY II: SEX

In this study, we are concerned with the differences between boys and girls in their identification with movie characters and their

<sup>5</sup> For girls, there was a tendency for the Billy identifiers to show the greatest advantage over Harry identifiers in the content which involved Billy's talking to Harry.



memory of movie content. Continuing with our investigation of the hypothesis that viewers tend to identify with characters similar to themselves in major social categories, we sought a movie which would provide fairly comparable opportunities for identifying with either a male or female character. Our prediction was that viewers would choose to identify with the like-sexed character, and that the movie content they learned would be in part a function of this choice.

The specific hypotheses Study II was designed to test were as follows:

1. Boy viewers tend to identify with the boy character and girl viewers with the girl character.

2. Viewers who identify with a character remember more of the words and actions of that character than viewers identifying with the opposite character. This prediction applies only to words and actions of a given character that are *not* directed toward the opposite character.

3. Viewers who identify with a character remember more of the stimuli directed toward that character than those viewers who do not identify with that character. (Stimuli include words and actions of secondary characters directed toward the identified-with character, as well as other environmental cues to which he is responding.)

4. There is no difference between the viewers who identify with the two primary characters in the amount of material they remember when the two characters are interacting with each other.

5. Boys remember more of the aggressive content of the movie than do girls.

### Method

*The movie.* The film was entitled *Tomboy*, a Hollywood feature made about 1940. The two primary characters were a boy and a girl, aged 14 or 15, who were pupils in a rural school. The story concerned a new girl in the community and a boy living with a cruel uncle, the developing interest of the boy and girl in each other, and the working out of the conflict between the boy and his uncle until the denouement with a happy ending.

*Subjects.* The Ss in this study were 7th-grade pupils in three public schools in the Boston metropolitan area. The schools served quite varied communities: in one school, over half the fathers' occupations were classified in Warner's classifications 1, 2, and 3; in a second school,  $\frac{2}{3}$  of the fathers' occupations were in Warner's

classifications 5, 6, and 7; in the third school, fathers' occupations were rather evenly distributed over the various classifications.

*Administration.* In each school, the movie was shown in an auditorium to all the 7th grade pupils. The E told the children that they were about to see an ordinary entertainment movie; he said that he wanted to find out something about what kinds of things children notice and enjoy in movies, and that he would ask them a few questions about their reactions to the movie later on. One week later the Es returned with questionnaires. The pupils remained in their classrooms and were tested one classroom at a time.

*Measuring instruments.* The measuring instruments were comprised of two parts: one designed to provide an indirect measure of identification; and the second designed to measure learning of the content of the movie. In addition, each S was asked to indicate his sex and his father's occupation.

As in Study I, we do not have a direct measure of the S's "identification" with a movie character, but employed an indirect measure intended to reveal aspects of each role's attractiveness for fantasy role playing. Since we found in the first study that the question "which character did you like the best" was not as useful in predicting the recall of movie content as the other items in our identification index, we omitted this question, and used the following questions: (a) "Which character is most like you?"; (b) "Which character would you like to be in real life?"; (c) "Which part would you like to play in the movie?" The Ss were given five choices—the two primary characters and three secondary characters—and were asked to indicate their first two choices to each question. Their answers were combined into an index, described below.

There are four kinds of content relevant to the hypotheses guiding this study: (a) behavior, including actions, speech, and feelings, of one primary character not directed toward the other primary character (designated "Alone" content); (b) behavior of one primary character directed toward the other primary character when they are in a scene together (designated "Interaction" content); (c) stimuli directed toward one of the primary characters but not by the other (designated "Stimulus" content); and (d) behavior of a primary character directed toward a secondary character in anger and/or with the purpose of injuring him either physically or psychologically (designated "Aggressive" content).

Three different procedures were used, on different groups of Ss, to measure the learning of content:

The first was a multiple-choice "recognition" test, similar to that used in Study I. This test was given to 291 students. In Study I there were disappointingly small differences in recall scores between viewers identifying with the two main characters. This fact suggested that our measuring instrument was relatively insensitive, and in Study II we therefore tried out two additional kinds of measuring devices in an effort to obtain sharper differences. One was a "recall" test, in which the Ss (47 in number) were required to write in their own answers to "open" questions. Another was a "reconstruction" test, given to 37 students, to be described more fully below.

## Results and Discussion

**Identification.** An "identification index" was computed on the basis of the answers to the three identification questions; this index indicated the degree to which one of the primary characters was seen as a more attractive role than the other. In computing index scores, each first choice was given a value of two, and each second choice a value of one. The index was equal to the score for the primary boy character minus the score for the primary girl character. Index scores showed very little cross-sex identification, as may be seen in Table 4. We can therefore test our hypotheses in terms of simple sex differences. All the following comparisons are thus in terms of boys versus girls.

**Learning.** As is so often the case at this age with verbal materials, the girls performed at a somewhat higher level of skill on all three measures of learning of movie content. For example, on the recognition test the girls got an average of 39.5 items correct, as compared with an average of 38.0 for the boys; this difference is significant at the .03 level. To allow for this difference and within-sex differences in total score, all the raw scores for each category of content were converted into percentage scores on the basis of the individual S's total number correct. The results for the three different types of measuring instruments are considered separately below.

**Recognition.** The recognition test was composed of 64 multiple-choice items about the content of the movie. There were four kinds of content relevant to our study, and some items in each of these content categories concerned the male leading character, some the female. The items were selected in terms of (and classified into) the following eight categories:

1. Boy Alone—boy character's actions and

words when the girl was not present. 11 items. (Example: "How did Steve react when his Uncle Mat told him he couldn't go to school anymore?")

2. Boy Interaction—boy character's actions and words directed toward the girl. 7 items. (Example: "When Pat and her father came by to see Uncle Mat on Sunday afternoon, what did Steve say to Pat when he first got out to the car?")

3. Boy Stimulus—stimuli directed toward the boy character but not by the girl. 11 items. (Example: "Steve was talking to Pat and her father on Sunday morning and suddenly said, 'Gee, it's late. I've got to run.' What reminded him he had to go?")

4. Boy Aggression. 7 items. (Example: "How many times did Steve hit Harry?")

5. Girl Alone. 7 items. (Example: "When Pat was getting lunch ready on the picnic, how did she show that she missed Steve?")

6. Girl Interaction. 7 items. (Example: "When Pat and her father stopped by at Steve's house after the picnic, what did Pat say to Steve?")

7. Girl Stimulus. 7 items. (Example: "When Pat went out the back door on Sunday morning looking for her father, what did she see?")

8. Girl Aggression. 7 items. (Example: "What did Pat do while the tramp was holding her?")

We predicted that boys would remember more aggressive content from the film than girls. This proved to be true. (See Table 5-A.) Furthermore, girls remembered more Girl Alone content, which was predicted, and more Girl Interaction content, which was *not* predicted. There were no significant differences on the other categories of content, although all were in the predicted direction.

**Recall.** The instrument used in the recall testing was composed of items that asked a question about the content of the movie and provided a space for the S's response. The items were selected from ones used in the recognition procedure, the stems being slightly reworded. There were five items from each of the eight categories of content. In scoring the recall questionnaire, two points were given for each essentially correct item, one point given for a partially correct item, and 0 points for each incorrect item. Points were added to

TABLE 4  
IDENTIFICATION WITH BOY AND GIRL MOVIE  
CHARACTERS BY SEX OF VIEWER

Identification Index Score	Choose Boy Character						No Choice	Choose Girl Character					
	6	5	4	3	2	1		6	5	4	3	2	1
Number of Boys	44	34	46	17	20	12	13	8	3	2	1	0	0
Number of Girls	0	1	1	1	0	6	7	7	13	26	23	37	53



TABLE 5  
 SEX DIFFERENCES IN MEMORY OF MOVIE CONTENT

	Content Categories							
	Boy Alone	Boy Stimulus	Boy Aggression	Boy Interaction	Girl Alone	Girl Stimulus	Girl Aggression	Girl Interaction
A: Recognition Test*								
Mean Boys' per cent score ( $N = 153$ )	18.0	16.6	8.6	10.7	11.5	13.7	9.7	11.3
Mean Girls' per cent score ( $N = 138$ )	17.9	16.4	7.4	10.7	12.3	14.2	9.5	12.1
$t$	.33	.53	3.50	.17	2.28	1.44	.56	2.12
$p$ †	n.s.	n.s.	.001	n.s.	.01	.10	n.s.	.05†
B: Recall Test*								
Mean Boys' per cent score ( $N = 27$ )	16.00	10.1	19.3	6.9	12.7	12.7	10.6	13.2
Mean Girls' per cent score ( $N = 20$ )	14.50	8.2	13.6	8.1	10.9	15.1	10.2	19.5
$t$	.78	1.36	3.98	.81	1.11	1.19	.25	3.15
$p$ †	n.s.	.10	.001	n.s.	n.s.	.12	n.s.	.005†
C: Reconstruction Test: Unit Classification*								
Mean Boys' per cent score ( $N = 20$ )	31.4	19.6	—	12.7	19.0	8.5	—	4.2
Mean Girls' per cent score ( $N = 17$ )	26.2	17.5	—	12.3	25.1	10.5	—	3.5
$t$	2.32	1.17		.20	3.38	1.70		.64
$p$ †	<.02	<.13		n.s.	<.001	<.05		n.s.

\* Each S's score is the number of items correct in a content category divided by the total number correct on the entire test.  
 † All  $p$ 's are one-tailed tests unless otherwise indicated.  
 ‡  $P$  is two-tailed: no direction was predicted.

provide a point score for each category of content and a total point score.

On the recall test, boys remembered more Boy Aggression content than did girls, and girls remembered more Girl Interaction content (see Table 5-B). Differences here are much larger than for the previously discussed recognition test. The differences obtained for Boy Aggression and Girl Interaction scores were at the same or higher level of statistical significance even though the test was shorter and only about  $\frac{1}{4}$  the number of Ss used.

**Reconstruction.** The reconstruction testing procedure involved the use of still photographic prints of scenes from the movie. The Ss were to respond to each stimulus picture by writing a paragraph or two telling what was happening at the moment of the picture, what had happened just before the picture, and what happened just after the picture. They were individually given nine  $3 \times 4\frac{1}{2}$  inch photographic prints, with four minutes to write about each. Three pictures depicted the primary boy character alone or with secondary characters; three pictures depicted the primary girl character alone or with secondary characters, and three pictures depicted the two primary characters, boy and girl, together.<sup>6</sup>

<sup>6</sup> Unfortunately, in printing, the pictures were reversed from right to left. The Ss did not report being confused about the scenes on this account, however.

The scoring of the resulting protocols was in terms of a "correct and relevant subject-predicate unit"; that is, each subject-predicate unit that actually occurred in the movie in the specific scene from which the picture was taken was counted. In addition, a score of one unit was given for the mention of material—no matter what the quantity in terms of subject-predicate units—that did not occur in the specific scene in question but was relevant to it either as a direct antecedent of the scene or a direct consequence of it. Material that actually occurred in the movie but was not directly relevant to the scene in question was not counted, nor was material that did not occur in the movie.

Each unit was classified if possible in one of the eight previously defined relevant categories. Less than two per cent of the content of the protocols could not be so placed. Aggressive material made up such a small proportion of the content of these protocols (about 3%) that it is not considered in the comparisons.

Half of the protocols were scored by two independent scorers. Interscorer reliabilities were computed by means of rank-order correlations. The reliability of the total scores was .90, while the reliability of the scores on the nine individual pictures ranged from .81 to .95 with a median of .86, and that of the scores for the six subcategories of content ranged from

.65 to .94 with a median of .80. Score-rescore reliabilities of a single scorer were somewhat higher.

The results obtained through this procedure may be analyzed in two ways: first, in terms of the individual unit classification, and secondly, in terms of the picture classification. In the first instance, boys and girls are compared on the basis of the score for each category of scorable units (Boy Alone, etc.), without regard for the classification of the picture for which the units were written. In the second instance, boys and girls are compared on the basis of the scores on each type of picture: Boy Alone, Girl Alone, and Interaction, without regard for the classification of each particular unit scored for that picture.

In the first analysis, boys obtained significantly higher scores on Boy Alone content; girls obtained significantly higher scores on Girl Alone and Girl Stimulus content; and there were no significant differences between boys and girls on Boy Stimulus, Boy Interaction, and Girl Interaction content (see Table 5-C). It should be noted that the differences obtained with the use of the reconstruction test are larger than those obtained on the recognition test.

In the second analysis, boys obtained statistically significant higher scores on the Boy Alone pictures; girls obtained significantly higher scores on the Girl Alone pictures; and there was no significant difference between the boys and girls on the Interactive pictures. (See Table 6.)

**Combined results.** Since the content categories for the three testing procedures were defined in essentially similar ways and the three procedures were used on independent sets of Ss, it is permissible to combine the probabilities obtained on the three procedures by the Stouffer technique (8). The results are shown in Table 7.

TABLE 6

SEX DIFFERENCES IN MEMORY FOR MOVIE CONTENT  
RECONSTRUCTION TEST: PICTURE CLASSIFICATION

	Classification of Picture		
	Boy Alone	Girl Alone	Interactive
Mean Boy per cent score ( <i>N</i> = 20)	35.3	27.0	38.3
Mean Girl per cent score ( <i>N</i> = 17)	30.4	33.9	35.6
<i>t</i>	2.43	3.52	1.21
<i>p</i> *	<.01	<.001	<.25

\* *P*'s are one-tailed tests except for the "Interactive" which is two-tailed.

Some of our original hypotheses are supported; others are not. We predicted that boys would remember more of the Boy Alone and Boy Stimulus content, and this proved to be the case. We also predicted that boys would remember more aggressive content, but we found that they remembered the aggressive content better than the girls only if the boy hero was the agent of aggression. As far as the girls are concerned, they remembered more Girl Alone and Girl Stimulus material, as predicted, but their greater memory for "girl" content did not extend to aggression, even when the girl was the agent of aggression. Furthermore, we predicted that there would be no difference between the sexes in their recollection of content in which the boy and girl were interacting, on the grounds that any action on the screen would constitute either an action of one's own character or a relevant cue for his behavior. We found, however, that girls *did* remember, better than boys, the words and actions of the girl character when they occurred in a context of interaction with the boy.

These findings suggest certain revisions of our hypotheses, still further in the direction

TABLE 7

COMBINED PROBABILITIES OF SEX DIFFERENCES OBTAINED IN THREE PROCEDURES OF STUDY II

	Content Categories							
	Boy Alone	Boy Stimulus	Boy Aggression	Boy Interaction	Girl Alone	Girl Stimulus	Girl Aggression	Girl Interaction
Recognition Test <i>Z</i> ( <i>N</i> = 47)	0.32	0.52	3.50	0.16	2.27	1.43	0.56	2.12
Recall Test <i>Z</i> ( <i>N</i> = 47)	0.77	1.34	3.68	-0.80	-1.09	1.17	0.25	2.90
Reconstruction Test <i>Z</i> ( <i>N</i> = 37)	2.22	1.15	—	0.20	3.08	1.65	—	-0.63
Combined <i>Z</i>	1.91	1.74	5.10	-0.25	2.46	2.45	0.57	2.54
Combined <i>p</i> (one-tail)	<.03	<.05	<.001	<.50	<.01	<.01	<.30	<.01



indicated in Study I. Memory for movie content appears to be influenced *both* by identification with a character in the movie and by the "need relevance" of a particular kind of content for the viewer. In cases where both identification and relevance are acting in the same direction, one should expect the greatest differences between boys and girls; where identification and relevance are acting in different directions, one should expect a minimum of difference between boys and girls; and where only identification is affecting the situation, the differences should be intermediate. If we assume that aggressive material is particularly relevant for boys, and boy-girl interactive material for girls,<sup>7</sup> the differences would be almost exactly of the order we find in Table 7, with Boy Aggression and Girl Interaction showing substantial differences, and Girl Aggression and Boy Interaction showing little or no differences.

#### SUMMARY

Our hypothesis was that the amount and kind of material a viewer learns from a movie is in part a function of his choice of a character in the movie with whom to identify himself. We expected that a viewer would tend to identify with a character similar to himself in major role characteristics (e.g., sex, age, or status) whenever the movie provided more than one sympathetically presented character. And we reasoned that the viewer, having chosen a character as the primary protagonist through whose eyes to experience the action of the movie, would remember more of that character's actions and words, and more of the stimuli to which that character was responding, than would the viewer who was identified with a different character.

In two studies we showed movies to 25 class rooms of seventh-grade children, and tested them a week later on their knowledge of movie content.

Using an indirect measure of identification,

<sup>7</sup> Our assumption about aggressive material being more relevant for boys, and romantic material for girls, is consonant with the findings of Dysinger & Ruckmick (2). These workers took PGR readings of children watching movies, and found a higher level of emotional response among boys to danger and conflict scenes, and among girls, to love and erotic scenes.

we found that viewers did indeed identify themselves with the like-sexed leading character, in viewing a movie which included both a strong male and strong female lead. With respect to similarity of social class, however, viewers were more likely to choose the protagonist whose social class corresponded with the viewer's *aspired* social class, rather than his current objective status.

As far as memory for movie content was concerned, we found that both the choice of a character with whom to identify and the "need relevance" of that character's actions were important. That is, viewers do tend to remember somewhat better the actions and words of the character with whom they identify, and the stimuli relevant for that character's actions, but the advantage in learning does not apply equally to *all* the actions of the viewer's chosen character. For boy viewers, aggressive content appears to be particularly relevant, for they remember aggressive content better than girls, *provided* its agent is the boy hero. For girls, boy-girl interactive content appears most relevant, and they remember this content better than boys whenever the girl heroine is the agent of the action.

#### REFERENCES

1. BLUMER, H., & HAUSER, P. M. *Movies, delinquency and crime*. New York: Macmillan, 1933.
2. DYSINGER, W. S., & RUCKMIK, C. A. *The emotional responses of children to the motion picture situation*. New York: Macmillan, 1933.
3. HAYES, K. J., & HAYES, C. Imitation in a home-raised chimpanzee. *J. comp. physiol. Psychol.*, 1952, **45**, 450-459.
4. HERBERT, J. J., & HARSH, C. M. Observational learning by cats. *J. comp. Psychol.*, 1944, **17**, 81-95.
5. MCKEE, J. P., & LEADER, FLORENCE B. The relationship of socio-economic status and aggression to the competitive behavior of preschool children. *Child Developm.*, 1955, **26**, 136-142.
6. MICHAEL, D. N., & MACCOBY, N. Factors influencing verbal learning from films under varying conditions of audience participation. *J. exp. Psychol.*, 1953, **46**, 411-418.
7. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155-178.
8. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954.

9. MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
10. ROSE, N., FAISON, E., & PODELL, J. Effects of rest pauses during a training film upon audience responses. *Amer. Psychologist*, 1953, **8**, 424. (Abstract)
11. WARDEN, C. J., FIELD, H. A., & KOCH, A. M. Imitative behavior in the Cebus and Rhesus monkeys. *J. genet. Psychol.*, 1940, **56**, 311-322.
12. WARDEN, C. J., & JACKSON, T. A. Imitative behavior in the Rhesus monkey. *J. genet. Psychol.*, 1935, **46**, 103-125.
13. WARNER, W. L., MEEKER, M., & EELLS, K. *Social class in America*. Chicago: Science Research Associates, 1949.
14. WRIGHT, G. O. Projection and displacement: A cross-cultural study of folk-tale aggression. *J. abnorm. soc. Psychol.*, 1954, **49**, 523-528.

Received September 17, 1956.



# BODY-IMAGE BOUNDARIES AND ADJUSTMENT TO POLIOMYELITIS

KENNETH E. WARE

*Baylor University College of Medicine*

*Southwestern Poliomyelitis Respiratory Center<sup>1</sup>*

SEYMOUR FISHER AND SIDNEY CLEVELAND

*Veterans Administration Hospital, Houston*

It has been a general observation that among patients afflicted with physically crippling illnesses or injuries, the psychological reaction to the trauma seems to bear little relationship to the extent or degree of the disability. Some patients with minimal physical incapacitation demonstrate a catastrophic reaction, a feeling of complete devastation, while others with extensive physical involvement are able to accept the prospects of permanent severe crippling and adjust to their handicap without evidence of marked personality disruption. Perhaps this difference in psychological reaction to a change in body structure may be accounted for, at least in part, by differences in the way individuals conceive of their own bodies. The concept of body image as a system of ideas and feelings which the individual has about his physical structure has had frequent application in the psychological literature on poliomyelitis and other disabling conditions. It has been pointed out (1, 5, 6, 9, 10) that disablement imposes upon the individual the necessity for a more or less drastic modification of his body image, an abrupt change in his perception of himself as a physical being. Difficulties in adjustment are often conceptualized as difficulties in incorporating the atrophied body, the facial disfigurement, or paralyzed limb into the basic, pre-existing body image. The ease or difficulty with which this integration takes place is an index of the success which one has with one aspect of the problem of readjustment.

Although body image is a concept often used in discussions of the psychological implications of physical handicap, there have been few attempts to derive systematic measurements relating to this phenomenon. Fisher and Cleveland (2, 3, 4), however, have described a method of obtaining measurements on one

dimension of body image. This dimension, called the body-image boundary, has to do with the degree to which the individual conceives of his body boundaries as definite and firm versus indefinite and weak. Its measurement is expressed in terms of what is called the Barrier score or the total number of a particular type of responses (called Barrier responses) produced by the individual on the Rorschach test. In addition to their published findings on normals and various psychosomatic symptom groups, Fisher and Cleveland have determined in unreported exploratory studies that degree of boundary definiteness is significantly linked with ability to tolerate stress in normal subjects and with facility of readjustment to body mutilation in individuals who have suffered limb amputation.

The concept of body-image boundary should have especially marked implications in a disease such as poliomyelitis in view of the fact that probably no other disease carries with it the actuality or threat of more severe physical incapacitation. It is conceivable that successful adaptation to such a disability may be positively related to the strength or firmness of the body-image boundary. The individual who already has a body image with a firm, definite, and protective boundary should be better able to withstand the threat incurred from physical insult, to incorporate the damage, and to re-establish the body-image boundary.

Within this framework, a study was designed to test the hypothesis that patients who achieve the better psychological adjustment to the physical disability imposed by poliomyelitis show, through a greater number of Barrier responses, evidence of a stronger, more definite body-image boundary than patients with a less satisfactory adjustment. Further, it was argued that the Barrier score, rather than reflecting the degree of the individual's physical incapacitation or the duration of his illness at the time of testing, is a

<sup>1</sup> Southwestern Poliomyelitis Respiratory Center, is aided by an annual grant from the National Foundation for Infantile Paralysis, Inc.

function of a more pervasive body-image factor in the personality which may dispose the individual toward a successful or unsuccessful adaptation to physical handicap.

### METHOD

The subjects (Ss) were 56 patients hospitalized in a poliomyelitis respiratory center which cares for patients in both the acute and postacute stages of illness. The group included all types of poliomyelitis cases: spinal, bulbar, and bulbo-spinal, both with and without respiratory muscle paralysis. It included cases with widely varying physical involvement, ranging from relatively mild to severe with a predominance of severe cases. It was regarded as representative of the usual patient population of the respiratory center. The group included 29 males and 27 females. The age range was from 15 to 43 years with a median of 26; education varied from 3 to 20 years with a median of 13; and Wechsler Verbal IQs (available on all but 4 patients) ranged from 72 to 145 with a median value of 114.<sup>2</sup>

In order to obtain the sample, files were examined on all patients, both current and discharged. Sixty Ss were found whose Rorschach protocols contained the minimum number of responses necessary for scoring the body-image variable. These patients were then evaluated in terms of adjustment to illness by procedures which will be described later. In the process of this evaluation, four patients were eliminated because of disagreement between judges as to adequacy of adjustment, leaving a total of 56 Ss.

The Rorschach test had been administered individually to each of these patients for general clinical assessment purposes prior to the inception of this study. In all cases the test had been given by an examiner who at the time had no knowledge of the method of determining body-image values from the Rorschach. Since the test had been given more or less routinely to adult patients, there was no indication of a selective factor in the administration.

Twenty-eight of the 56 Ss had been tested by the senior author and at the time of the study were either still in the hospital or had only recently been discharged. The remaining 28 Ss had been tested and subsequently discharged prior to the senior writer's employment at the respiratory center. These two groups, which were treated separately as independent samples, are referred to as Groups I and II, respectively. Adjustment ratings were made in somewhat different fashion for the two groups and involved different sets of judges. It was of interest to determine whether a relationship between body-image data and

adaptation to disability occurred in both samples. If so, it would suggest a degree of stability in the relationship.

Groups I and II were each subdivided into two categories relating to the adequacy of adaptation to illness. The method used was that of gross clinical evaluation. In the case of patients in Group I, three staff physicians who knew the patients well and a psychologist (the present senior author), made independent gross evaluations of the 32 Ss on whom Rorschach records of sufficient length were available. These evaluations were made prior to inspection of the Rorschach protocols with reference to the body-image variable. Judges were asked to divide this group of patients into two categories of approximately equal number, placing those who had shown the most satisfactory adaptation to their illness in one group, and those with the least satisfactory adaptation in another. No attempt was made to define for the judges the criteria for satisfactory or unsatisfactory adjustment. They were simply told to use their global clinical impression of the patient's behavior during his period of hospitalization in allocating him to one or the other of the two categories. In 28 out of the 32 cases, at least three out of the four judges agreed as to placement of the patient. These 28 individuals composed Group I. The "satisfactory" and "unsatisfactory" categories thus identified included 15 and 13 patients, respectively.

Because the 28 patients in Group II had been discharged for some time, it was felt that staff members would not remember them well enough to make an evaluation of their adjustment while in the hospital. However, an index of adjustment was already available in the records of these patients. Soon after the patient's discharge, the social worker at the Center had rated on a 5-point scale the adequacy of the patient's overall psychological adjustment to his illness during the course of his hospitalization. This evaluation in a sense represented the pooled judgments of the entire staff. It was essentially an integration by the social worker of her own impressions and the opinions expressed by various members of the staff during a number of conferences on the patient and in the course of informal contacts between the social worker and other staff members while the patient was in the respiratory center. This entry, known as "Patient's Response to Disease (PRD)," is one dimension of the "Social Service Profile,"<sup>3</sup> which has been in use at the Center for a number of years. It is completed on each patient by the social worker as a routine procedure. For purposes of illustration, the definitions of the two extremes of the PRD scale are shown as they appear in the Profile:

*PRD 0.* Degree of anxiety and apprehension present does not seem out of proportion to the neurological involvement; cooperates to the extent of his ability. His responses to illness and hospitalization do not interfere with his acceptance of treatment. Can utilize his residual abilities rather than withdrawing and becoming completely dependent on

<sup>2</sup> This high median intelligence level should not be taken as an indication of selective sampling. The Wechsler-Bellevue Verbal Scale has been given to more than a hundred adult patients at the respiratory center by various examiners during the past three years, and the median IQ for these data corresponds very closely with the median IQ of 114 reported for the Ss used in this study. Apparently adult poliomyelitis patients, at least those treated in this center, are as a group above average in intelligence.

<sup>3</sup> The "Social Service Profile" was designed by Maurine B. Mitchell, Director of Medical Social Work, Southwestern Poliomyelitis Respiratory Center and Wolff Home Poliomyelitis Rehabilitation Unit.



others; is considerate of the rights and privileges of others (within his physical limitations).

**PRD 4.** Physiologically unjustified amount of extreme anxiety; insecurity and apprehension prevent cooperation; may result in patient becoming either fearful, belligerent, hostile, or rebellious; fails in making acceptable adjustment to hospitalization and separation from family; complete obstruction of treatment program.

For the patients in Group II, dichotomizing the distribution of PRD scores into categories of 0-1 and 2-4 yielded two groups of 14 patients each, which were designated as "satisfactory" and "unsatisfactory" adjustment groups.

Descriptive data on the total sample and on the four subgroups are presented in Table 1.

The method used for scoring body-image boundary definiteness from the Rorschach has been described in detail elsewhere (2). The basic assumption underlying the scoring procedure is that the definiteness of the individual's body-image boundaries is reflected in the manner in which he describes the peripheries of his ink blot percepts. Definite boundaries are linked with descriptions which emphasize the protective, decorative, covering, containing, and concealing attributes of percept peripheries. The following are examples of Rorschach responses which exemplify the attributing of such special qualities to the periphery: "cave with rocky wall," "mummy all wrapped up," "man in suit of armor," "fort with high walls," "woman with fancy decorative dress." The number of responses of this sort given by an individual in his Rorschach record is taken as an index or score of boundary definiteness and is referred to as the Barrier score. It has been shown that the Barrier score can be determined by scorers with a high degree of objectivity (2).

Because the Barrier score is probably correlated with the total number of Rorschach responses produced, it was necessary to control variations in response total as much as possible. This was done by using only records with at least 15 responses and by reducing to 25 responses all records with more than 25. In the case of records with R greater than 25, this procedure involved using only the first three responses on each of the first five Rorschach cards and the first two re-

sponses on each of the last five cards. When there was less than the set quota of responses on any given card, it was compensated for by taking an extra response on the next card in the sequence for which there was more than the set quota.

After the records had been scored for Barrier responses, using this method of controlling R, the group of 56 Ss was dichotomized at the median of the Barrier score distribution to yield a high- and a low-Barrier group. It was found that these two groups each had a response total of 25. Although the range of responses was 15-25, actually only a few cases in each group fell below 25, so that the median in both cases corresponded to the upper limit of possible response totals. Since the medians were the same, it may be concluded that there was no significant difference between high- and low-Barrier Ss with respect to R.

Intelligence was a second factor to be considered in addition to R as possibly affecting the production of Barrier scores. The median IQ for the high-Barrier group was 114 and for the low-Barrier group 113.5. Accordingly, it was assumed that intelligence did not contribute to the difference in Barrier scores between the two groups.

## RESULTS

Testing the hypothesis, which asserted a relationship between body-image boundary and adjustment to illness, required a comparison of "satisfactory" and "unsatisfactory" groups with respect to the frequency of Barrier scores produced. As shown in Table 2, the median Barrier score for patients in subgroups characterized by the more satisfactory adjustment seemed to exceed that of patients who made less satisfactory adjustments. Within each of the two samples (Groups I and II), chi square was used as the test of significance in comparing "satisfactory" and "unsatisfactory" patient groups with regard to the number of Ss above and below the median of the Barrier score distribution. The Barrier score frequency

TABLE 1  
CHARACTERISTICS OF SUBJECTS CLASSIFIED AS HAVING "SATISFACTORY" AND "UNSATISFACTORY" ADJUSTMENT  
AND OF THE TOTAL GROUP

Group and Adjustment Category	N	Sex		Age		Education		W-B Verbal IQ		Months Ill at Testing*	
		M	F	Mdn	Range	Mdn	Range	Mdn	Range	Mdn	Range
Group I											
Satisfactory	15	6	9	26	20-34	12	9-17	113	76-129	3	1-14
Unsatisfactory	13	6	7	26	19-43	13	9-17	113	72-132	8	1-17
Group II											
Satisfactory	14	10	4	25	18-38	15	3-17	115	80-133	7	1-14
Unsatisfactory	14	7	7	27	15-33	12	6-20	119	81-145	12	6-18
Total group	56	29	27	26	15-43	13	3-20	114	72-145	7	1-18

\* By Rorschach test.

TABLE 2

MEDIAN AND RANGE OF BARRIER RESPONSES FOR SUBJECTS CLASSIFIED AS HAVING "SATISFACTORY" AND "UNSATISFACTORY" ADJUSTMENT AND FOR THE TOTAL GROUP

Group and Category	Median	Range
Group I		
Satisfactory	4	1-7
Unsatisfactory	2	1-7
Group II		
Satisfactory	4.5	0-8
Unsatisfactory	2.5	1-5
Total group	3	0-8

was distributed in categories of 0-3 and 4 or more responses. Correction for continuity was used in both analyses. Results of these procedures are shown in Table 3.

Examination of Table 3 reveals that the incidence of Barrier scores in both Groups I and II was in the predicted direction with patients in the "satisfactory" categories showing the greater frequency. In the case of Group I, the difference in Barrier score frequency between the "satisfactory" and "unsatisfactory" categories was significant at the .11 level and for Group II at the .06 level of confidence. Since the assumption of independence of experimental outcomes was satisfied, the two probabilities were combined to yield a composite chi square (8, pp. 46-47). The probability associated with this chi square would be the aggregate probability of obtaining the given set of results from two independent tests of the relationship between the adjustment and body-image variables. The composite chi square was found to be 10.06. With four degrees of freedom (twice the number of probabilities involved), this value for chi square falls between the five and the two per cent levels of significance. Thus, although results of the two independent analyses did not prove to be strictly significant, the collective results support the hypothesis. In other words, it would appear that people who show evidence of definite body-image boundaries, in comparison with those having indefinite boundaries, tend to adjust significantly better to the experience of physical disability resulting from poliomyelitis.

The question remains of whether the Barrier score reflects a relatively well established body-image factor in the personality, or is simply

TABLE 3

COMPARISON OF PATIENTS CLASSIFIED AS HAVING "SATISFACTORY" AND "UNSATISFACTORY" ADJUSTMENT ON FREQUENCY OF BARRIER RESPONSES

Group	Satisfactory		Unsatisfactory		$\chi^2$	$p$
	Above Mdn	Below Mdn	Above Mdn	Below Mdn		
I	9	6	3	10	2.58	.11
II	9	5	3	11	3.65	.06

a function of the degree of physical incapacitation or duration of illness at the time of testing. Three steps were taken in examining this problem. The chief physician was asked to pick from the total group the 15 patients whom he considered to be the most incapacitated from the physical standpoint, and the 15 patients with the least physical involvement. This procedure was used because accurate ratings on degree of physical disability could be made only for extremes. Chi square was used as before in comparing the two groups with respect to the number of individuals with Barrier response totals above and below the median of three. The results showed no difference in Barrier scores for groups characterized by different degrees of physical involvement.

As a second step, the relationship of Barrier scores to duration of illness at time of testing was studied in the total group by means of chi square. It was found that length of illness at time of testing for high-Barrier Ss did not differ significantly from the duration of illness at time tested for low-Barrier Ss. Median durations for the high- and low-Barrier groups were 7 and 8 months, respectively. It should also be noted that Fisher and Cleveland, in a previous study (2), found no relationship between Barrier score and duration of illness in a group with rheumatoid arthritis and a group with neurodermatoses.

These observations are consistent with the argument advanced here, but it is evident from Table 1 that both of the groups rated "unsatisfactory" had been ill for a considerably longer period of time when tested than those rated "satisfactory" on adjustment. This raised the question of whether the significant relationship between Barrier score and adjustment might be contaminated by the variable of duration of illness. In order to control the



possible influence of this variable, a matched pair technique was utilized as a third step in the analysis.

It was possible to select from the total sample a group of 15 "satisfactory" Ss, each of whom was closely matched (within one month) with respect to duration of illness by a patient among the "unsatisfactory" Ss. The median duration at time of testing was 7.5 months, and the range was 1 to 14 months for each of the two groups thus obtained. Chi square, corrected for continuity, was used to compare the two groups with regard to the number of Ss above and below the Barrier score median. With length of illness at time tested held constant, high-Barrier individuals still fell more frequently in the "satisfactory" category (.06 level of confidence). This result did not quite reach the conventional criterion for statistical significance, but it was regarded as a fair indication that the Barrier score is not (or at least not primarily) a function of how long the patient had been ill when he was tested.

It might be noted, incidentally, in speculating as to why the more poorly adjusted patient was tested at a later date in his hospitalization than someone with a better adaptation to his illness, that the most plausible explanation is that the patient who made a poor adjustment was simply a more disagreeable and more difficult individual to work with and was likely to be resistant to testing. There was probably a tendency to delay and procrastinate in seeing this patient, especially since it was known in most cases that he would be available for testing over a period of several months of hospitalization.

#### DISCUSSION

The results confirm the major hypothesis involved in this study and also furnish evidence that would seem to make the argument advanced for the nature of the Barrier score a tenable point of view. The correlation found to exist between the body-image variable and adjustment seems to indicate that the extent to which the individual conceives of his body exterior as having a defensive, shielding, barrier-like quality plays some part in the total constellation of forces affecting the adequacy of the individual's adaptation to

physical crippling. Apparently, the individual whose body image includes these qualities is better able to cope with the problems imposed by his handicap than is the person whose body concept lacks these characteristics.

It is readily evident, in the light of the considerable degree of overlap in Barrier score frequently between "satisfactory" and "unsatisfactory" groups, that the body-image variable, at least as presently defined, possesses little diagnostic or predictive efficiency. Nevertheless, the obtained relationships suggest that the concepts involved are theoretically fruitful and merit further inquiry.

In general, the Barrier score seems to reflect some kind of a body-image factor in the personality tending to dispose the individual toward a favorable or unfavorable adjustment to physical handicap. This study, of course, made no direct test of this contention, and such a test would be difficult to make. However, if contemporary circumstances affect the production of Barrier responses, it would seem that the duration and severity of illness would have influenced the Barrier score in the group studied. Since the results indicate that the Barrier response is probably independent of these variables, the overall observations at least seem congruent with the idea that the Barrier score represents a personality factor of a fairly basic nature.

The correlation found between adjustment and production of Barrier responses is quite consistent with results obtained by Fisher and Cleveland in a study of relationships between the Barrier score and certain personality variables (3). They found that the Barrier score in normals is related to such characteristics as greater participation in active sports, higher *n* Achievement scores, the production of TAT themes involving identification with a more definite image of the parents, and more emphasis on task completion. They concluded that emphasis on the Barrier aspect of body-image boundary is "a reflection of a style of life based on an unusually strong definition of self-identity and on objective self-expression aimed at establishing a stable, controlling relationship to the environment." This implies a degree of self-confidence that should serve as a bulwark against a threat to self-integration. The individual possessing these characteristics

obviously should be better able to absorb the threat to the self involved in physical loss or crippling and to integrate this loss into a revised self concept with the minimum of psychic trauma.

It should be noted that the body-image score has already proved meaningful in another crippled group with respect to the incorporation of loss. In an unpublished study by Fisher and Cleveland on data supplied by Haber (7), it was found that high-Barrier amputees dealt more realistically with phantom limb experiences than did low-Barrier amputees. One might conceptualize this as indicating that high-Barrier people were better able to integrate the body mutilation into their body-image schema.

These experimental observations lend support to the theoretical significance generally assumed for the body image as a factor in an individual's adaptation to physical handicap. It would seem, therefore, that the study of the individual's attitude toward his body is a promising way of approaching the problem of adjustment to disability. Meaningful relationships have been demonstrated between adjustment and the boundary aspect of body image. Perhaps it would be rewarding to extend investigations concerning physical handicap into other dimensions of the body concept.

#### SUMMARY

This study has shown, in a group of 56 poliomyelitis patients, a relationship between psychological adjustment to physical handicap and a measure of the boundary aspect of body image. This measure, derived from the Rorschach protocol, is known as the Barrier score. Apparently, individuals who conceive of their body boundaries as possessing defen-

sive, armoring, barrier-like qualities are significantly better able to adapt to the circumstances surrounding physical loss than are those individuals whose body concept does not contain these features. The investigation of other dimensions of body image is suggested as a promising approach to the study of adjustment to handicap.

#### REFERENCES

1. BERGER, S. Paraplegia. In J. F. Garrett (Ed.), *Psychological aspects of physical disability*. Department of Health, Education, and Welfare, Office of Vocational Rehabilitation. Washington: U. S. Government Printing Office. Pp. 46-59.
2. FISHER, S., & CLEVELAND, S. E. The role of body image in psychosomatic symptom choice. *Psychol. Monogr.*, 1955, **69**, No. 17 (Whole No. 402).
3. FISHER, S., & CLEVELAND, S. E. Body-image boundaries and style of life. *J. abnorm. soc. Psychol.*, 1956, **52**, 373-379.
4. FISHER, S., & CLEVELAND, S. E. Relationship of body-image boundaries to memory for completed and incomplete tasks. *J. Psychol.*, 1956, **42**, 35-41.
5. GLUD, E., & BLANE, H. T. Body-image changes in patients with respiratory poliomyelitis. *Nerv. Child*, 1956, **11**, No. 2, 25-39.
6. GRAYSON, M. Concept of "acceptance" in physical rehabilitation. *J. A. M. A.*, 1951, **145**, 893-896.
7. HABER, W. B. Effects of loss of limb on sensory functions. *J. Psychol.*, 1955, **40**, 115-123.
8. LINDQUIST, E. F. *Statistical analysis in educational research*. New York: Houghton Mifflin, 1940.
9. MENNINGER, K. Psychiatric aspects of physical disability. In J. F. Garrett (Ed.), *Psychological aspects of physical disability*. Department of Health, Education, and Welfare, Office of Vocational Rehabilitation. Washington: U. S. Government Printing Office. Pp. 8-17.
10. PRUGH, D. G., & TAGIURI, C. K. Emotional aspects of the respirator care of patients with poliomyelitis. *Psychosom. Med.*, 1954, **16**, 104-128.

Received September 19, 1956.



# THE MANIFEST ANXIETY SCALE AS A MEASURE OF DRIVE

ROBERT E. SILVERMAN

*New York University*

THE Taylor Manifest Anxiety Scale (MAS) (16, 17) was designed to provide a technique for distinguishing drive level in humans in order to test aspects of Hullian learning theory. It was assumed that the scale measured differences in emotional responsiveness which contribute to the total effective drive state. According to Taylor (17), the title of the test should not be construed as suggesting special clinical validity since the scale was not designed for clinical purposes. In addition, Taylor and her associates have not been concerned with the relationship of anxiety per se to performance.

In spite of the original purpose of the MAS, many investigators have sought to relate the scale to clinical criteria of anxiety (3, 6, 8, 12); others have persisted in attempting to relate anxiety to performance variables (1, 2, 5, 11, 15, 18). Thus, many investigations owe their origins to Taylor's recent work and to Hullian theory. If these many studies have not furnished clear answers, they have served an heuristic purpose in highlighting the following questions: (a) What are the relationships between drive level and performance? (b) Does anxiety intensify effective drive level? (c) Does the Taylor MAS identify anxiety? (d) Does the Taylor MAS identify differences in drive level?

The first question continues to be extensively treated by students of Hullian theory, while the second and third questions await an unequivocal criterion of anxiety. The purpose of the present study was to consider the fourth question directly since the assumed relationship between the MAS and drive has been the focal point of a variety of recent investigations.

If one is willing to argue that drive may be defined solely in terms of the MAS, then the question becomes superfluous. However, it would seem that the intention of the Iowa group was to relate general drive variables to performance, not simply MAS scores to performance. In this light, the question becomes: What are the relationships between the MAS and measures of emotional responsiveness? If

the MAS is a measure of emotionality, then, according to Taylor, it should be related either to a "chronic emotional state" in which subjects (Ss) demonstrate high levels of emotionality or to differences in the potentiality for arousal, manifested by different degrees of emotional reactions to novel or threatening situations.

Palmar skin conductance and conductance changes were employed as measures of emotional responsiveness in the research to be reported. This technique, recently described by Schlosberg (14), has the multiple advantages of simplicity, reliability, and accuracy in estimating general states of autonomic nervous system activity.<sup>1</sup>

Because two recent investigations have failed to demonstrate a linear relationship between the MAS and the GSR or between the MAS and palmar perspiration technique (1, 9), steps were taken to explore the conditions that might obscure such relationships by employing two psychometric inventories. The two inventories were the MAS and a forced-choice version of the MAS designed by Heineman (7). The forced-choice inventory was included to provide some control over one of the known sources of error in the MAS, the influence of social desirability variables (7, 12).

## METHOD

### *Subjects*

The Ss were 85 male undergraduate students from the elementary psychology course at the University College of Arts and Sciences at New York University. The Ss were administered an inventory containing the Taylor MAS and the K scale from the Minnesota Multiphasic Personality Inventory as a part of the weekly small group discussions held in conjunction with the lecture course. The skin conductance readings were taken in a laboratory setting bearing no physical, temporal, or otherwise obvious relationship to the tests. In all, 66 Ss participated in the MAS and the laboratory

<sup>1</sup> A study by the author, recently completed, indicates week-to-week reliability coefficients for the conductance measures ranging from .68 to .74. In addition, conductance changes were found to be systematically related to shock stress.

phases. Six weeks after the MAS was administered, the Heineman forced-choice version of the original Taylor scale (7) was given again in the discussion meetings, this time to 77 of the original Ss.<sup>2</sup> The attrition of Ss was due mainly to class absences and failure to keep appointments.

### Apparatus and Procedure

The apparatus for measuring skin conductance consisted of a 50 microampere meter in series with two quarter-sized silver-silver, chloride-saline paste electrodes. This is essentially the unit outlined by Schlosberg (14). The circuit was run by a single one-and-one-half volt flashlight battery which, with the aid of a potential divider and a calibrated resistor, permitted absolute conductance to be read directly. In order to insure consistent electrode placement from S to S, and to maintain optimal contact during the experimental session, the electrodes were attached to S by having him grasp a cylindrical two-inch wooden bar on which the electrodes were seated in rubber cups. A second padded bar was then firmly clamped over each hand. Thus, S was in a kind of hand stock which permitted finger movements without disturbing the electrode contact.

Each S was introduced into the experimental setting by being placed in the stock and instructed to relax and rest while the experimenter (E) recorded his skin temperature. (This subterfuge was employed in order to obviate any concern on S's part with electric currents and possible shock.) During this rest period of ten minutes, E recorded two conductance levels every 30 seconds, one from each electrode. This was done to insure the obtaining of the optimum conductance reading. Only the highest reading from each pair of readings was used in computing the mean conductance level for a given S.

Following this rest period, all Ss were instructed to perform a simple series of tasks consisting of 25 simple addition problems presented on a small screen directly in front of them. One group of Ss ( $N = 36$ ), selected at random, was assigned to the shock-threat condition. They were fitted with an arm band containing two electrodes and were instructed that,

During the next few minutes you are going to be shown some numbers on this screen. You are to add the numbers as quickly as possible... but also as accurately as you can. If you are not performing well, you will receive a shock... something like this. (A sample shock from an Applegate Constant Current Stimulator set to deliver 2 milliamperes was given here.) The no-shock Ss ( $N = 30$ ) were given the same task instructions but were not fitted with the electrodes and were not threatened with shock.

The E recorded two conductance readings immediately preceding the presentation of each problem. The entire problem series lasted 12½ minutes.

### RESULTS AND DISCUSSION

The mean conductance scores for the rest periods were transformed into logarithmic

<sup>2</sup> Heineman's FC-1 was used, scored according to Key 2 as described in the original report (7).

units for the purpose of statistical comparisons.<sup>3</sup> The shock and no-shock task conditions provided measures of arousal which were computed in terms of the changes in log conductance from the rest periods to the task periods.

The correlations of the MAS scale and the forced-choice scale (FC) with the log conductance readings are shown in Table 1. None of the MAS correlations approaches significance. However, the FC scale does show a reliable if slight relationship with the rest period readings ( $r = .24$ ) and a clear negative relationship with the conductance changes under the threat of shock ( $r = -.50$ ). This negative relationship indicates that high scores on the FC scale are associated with low degrees of arousal.

An explanation for the fact that the two scales, the MAS and the FC, consist of similar items but do not correlate similarly with the conductance measures, may be found in Table 2 where the relationships among the MAS, the FC, and the K scales are shown. The K scale was designed to be a measure of "test-taking attitude" (10). High K scores are thought to be associated with a type of defensiveness against psychological weakness, while low K scores may indicate excessive self-criticism. Thus, the correlation of  $-.55$  between the MAS and the K scale may be an indication that some low MAS scores were associated with defensiveness and some high MAS scores were simply reflecting a too candid or critical attitude. The FC scale, although correlating reliably with the MAS, did not show a relationship with the K scale.

These findings suggest that the failure to demonstrate a correlation between the MAS and the conductance measures may have been due to the spurious nature of some of the MAS scores. When steps were taken to minimize the effects of the test-taking attitude by employing a forced-choice technique, a relationship was found between the questionnaire and the physiological measure. Thus, an inventory like the Heineman scale appears to be a more sensitive measure of emotional responsiveness than the MAS.

It was observed above that conductance

<sup>3</sup> In addition to being normally distributed, log conductance changes have been found to correlate well with measures of palmar perspiration. Mowrer *et al.* (13) reported a relationship between palmar perspiration, measured by a colorimetric technique, and measures of "psychological" tension during psychotherapy.



TABLE 1  
CORRELATIONS BETWEEN THE TWO SCALES AND LOG  
CONDUCTANCE READINGS

	MAS		FC	
	r	SE	r	SE
Conductance: rest	.07	.11	.24*	.11
Conductance change: no shock	-.14	.18	.12	.20
Conductance change: shock	-.17	.16	-.50**	.14

\* Significant at the .05 level.

\*\* Significant at the .01 level.

TABLE 2  
INTERCORRELATIONS AMONG THE THREE SCALES

	FC	K
MAS	.50** (N = 69)	-.55** (N = 85)
FC	—	-.07 (N = 69)

\*\* Significant at the .01 level.

change is inversely related to the FC scale under the shock threat condition. This suggests that persons defined as anxious by the scale are less aroused by a threatening situation. These findings are in direct contradiction of the hypothesis that high levels of emotional responsiveness are associated with increased potentialities for arousal. Nevertheless, an interesting and not unfamiliar relationship is indicated which points up possible shortcomings in the conception of anxiety as a drive. Silverman and Blitz (15), using rote learning measures, found that high anxious Ss (defined by MAS) were less affected by threat conditions than their low anxious controls. Deese, Lazarus, and Keenan (5) similarly reported that the rote learning of high anxious (MAS) Ss was not reliably affected by shock stress. In both of these studies, the dependent variable, rote learning, did not offer any clue in regard to other behaviors possibly associated with threat and related to rote learning.

Yet it is unlikely that anxious persons are generally unresponsive to threat situations. It may be that the threat conditions of the laboratory are relatively benign by comparison to the personal doubts and fears which anxious persons habitually carry with them. Expressed in terms of this study, one might say that the tensions engendered by the laboratory threat

did not raise an already high state of emotionality. A more general conception is that increases in arousal depend upon the already existing state of tension. This conjecture is supported by the fact that in the present study the amount of conductance change was a function of conductance level during rest. The product-moment correlation between rest level and change was .63. Darrow (4) also found that the size of the autonomic reaction (skin resistance) is a function of the amount of "available energy" which a given individual has at his disposal. Thus, it would seem that persons whose tension levels are high initially do not show much arousal and, hence, are unable to cope with novel situations which demand increases in arousal.

#### SUMMARY

This study was concerned with the relationship between the Taylor MAS and skin conductance. A group of Ss was given the MAS and six weeks later the Heinman forced-choice version of the scale. Skin conductance measures were obtained for each S under two conditions: (a) a rest period, and (b) a task period involving shock threat for some Ss, and no shock threat for others.

The results indicated that the forced-choice scale was positively related to the readings taken under rest condition, and negatively related to changes in conductance obtained under the threat of shock. The MAS did not correlate with any of the conductance readings.

These findings were explained by considering the high correlation between the K scale and the MAS and by noting that anxiety has been found by other investigators to be associated with low arousal.

#### REFERENCES

1. BEAM, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.*, 1955, 51, 543-551.
2. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, 47, 615-623.
3. BUSS, A. H., WIENER, M., DURKEE, A., & BAER, M. The measurement of anxiety in clinical situations. *J. consult. Psychol.*, 1955, 19, 125-129.
4. DARROW, C. W. The equation of the galvanic skin reflex curve: I. The dynamics of reaction in

- relation to excitation background. *J. gen. Psychol.*, 1937, **16**, 285-309.
5. DEESE, J., LAZARUS, R. S., & KEENAN, J. Anxiety, anxiety-reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55-60.
  6. GLEESER, GOLDINE, & ULETT, G. The Saslow-Screening test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8**, 279-283.
  7. HEINEMAN, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447-454.
  8. KENDALL, E. The validity of Taylor's Manifest Anxiety Scale. *J. consult. Psychol.*, 1954, **18**, 429-432.
  9. LOTSOFF, E. J., & DOWNING, W. L. Two measures of anxiety. *J. consult. Psychol.*, 1956, **20**, 170.
  10. MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The K scale. *J. consult. Psychol.*, 1948, **12**, 20-31.
  11. MALMO, R. S., & AMSEL, A. Anxiety-produced interference in serial rote learning with observations on rote learning after partial frontal lobectomy. *J. exp. Psychol.*, 1948, **38**, 89-101.
  12. MATARAZZO, J. D. MMPI validity scores as a function of increasing levels of anxiety. *J. consult. Psychol.*, 1955, **19**, 213-217.
  13. MOWRER, O. H., LIGHT, B. H., LURIA, X., & ZELENY, M. P. Tension changes during psychotherapy with special references to resistance. In O. H. Mowrer (Ed.), *Psychotherapy*. New York: Ronald Press, 1953. Pp. 546-640.
  14. SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81-88.
  15. SILVERMAN, R. E., & BLITZ, B. Learning and two kinds of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 301-303.
  16. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
  17. TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303-320.
  18. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and flexibility of the PGR. *J. Psychol.*, 1947, **23**, 83-91.

Received September 19, 1956.



# THE PSYCHOLOGICAL MEANING OF ACQUIESCENCE SET FOR AUTHORITARIANISM<sup>1</sup>

N. L. GAGE, GEORGE S. LEAVITT,<sup>2</sup> AND GEORGE C. STONE<sup>3</sup>

*Bureau of Educational Research, University of Illinois*

**T**HIS paper is directed against the argument that the California F (predisposition to fascism) scale (1) and similar inventories have their validity lowered by their loading with acquiescence set. The argument runs as follows: (a) the F scale consists entirely of "negative" statements, i.e., statements that, when agreed with, reflect authoritarian attitudes; (b) the tendency to agree with ambiguous and unstructured statements has high internal consistency and generality over tests of markedly different content (7); (c) the F scale, and other inventories of attitudes about social issues and interpersonal relationships which employ agree-disagree items, are significantly loaded with irrelevant variance in acquiescence tendency.

Thus, Cohn (5) found the F scale significantly correlated (.41) with his Plus scale; the latter measures the number of times one responds "true" to 33 MMPI items dealing, according to Cohn, with nothing in particular. Similarly, Bass (4) found the F scale correlated positively with an acquiescence factor obtained by factor analysis.

## A CORROBORATIVE STUDY

The present authors have obtained similar results in an independent study. We used Cohn's Plus Scale, the F scale, the Minnesota Teacher Attitude Inventory (MTAI) (6), and an especially constructed "Information-True" test of acquiescence. These were administered to 118 graduate students in education at the University of Illinois.

The Information-True test consisted of 50 "difficult" true-false information items. These items had been written and found through pretesting to be sufficiently difficult and obscure to elicit approximately a 50-50 split of "true" and "false" responses. The difficult items were

mixed with 40 "easy" items; the latter were written and found through pretesting to be at about the 75 per cent level of difficulty. (The corrected split-half reliability of the number-correct scores on the easy items was .35; the latter score served as a crude measure of information about history, science, civics, Gallup Poll data, geography, and the like.)

The corrected split-half reliability of the acquiescence score, number of "true" responses on the difficult information items, was .68. This  $r$  indicated that the test was successful in eliciting reliable individual differences in tendency to respond "true," or acquiescence set. That difficult and ambiguous items are required to elicit the acquiescence set is again demonstrated by the fact that the reliability of the acquiescence score on the easy information items was only .09.

Table 1 shows the correlations among the Cohn Plus Scale, the Information-True score, the MTAI, and the F scale. These  $r$ s corroborate previous findings. The significant positive correlation (.19) between our Information-True score for acquiescence and Cohn's Plus Scale shows that acquiescence tendency is indeed general over considerably different types of test content. The correlation probably would have been higher if Cohn's Plus Scale were not concerned with another variable, namely, tendency to self-disparagement. That is, on most of Cohn's 33 MMPI items, the "true" response seems also to be the unfavorable description of oneself. This impression was supported by the judgments of two graduate students and two clerical workers who were asked to indicate whether "true" or "false" would constitute an unfavorable self-description in response to the 33 items. Their independent judgments agreed unanimously that "true" was the self-disparaging response to almost all the items.

Table 1 also shows that the Plus scale and the Information-True score correlate negatively ( $-.29$  and  $-.26$ , respectively) with the MTAI. These negative  $r$ s were expected on the

<sup>1</sup> This investigation was supported in part by a research grant (M-650) from the Institute of Mental Health, National Institutes of Health, Public Health Service.

<sup>2</sup> Now at Fresno State College.

<sup>3</sup> Now at G. D. Searle & Co., Skokie, Illinois.

TABLE 1  
CORRELATIONS AMONG ACQUIESCENCE, MTAI, AND  
F SCALE  
( $N = 118$ )

Measure	Mean	SD	$r_{tt}$	$r$ 's		
				(2)	(3)	(4)
1. Acquiescence: Plus Scale	12.7	5.1	.76†	.19*	-.29**	.44***
2. Acquiescence: Information-True	27.2	5.7	.68†		-.26**	.35***
3. MTAI	53.8	32.4	.86†			-.62***
4. F scale	49.8	23.8	.83†			

\* Significant at the .05 level.

\*\* Significant at the .01 level.

\*\*\* Significant at the .001 level.

† Corrected split-half.

‡ Kuder-Richardson formula 20.

grounds that (a) the acquiescence tendency is general over highly diversified types of test content, and (b) the MTAI scoring weights are negatively correlated with degree of agreement to its items. The latter point is apparent from the published scoring keys for the MTAI. Of the 274 +1 scoring weights in the MTAI key, 60 per cent are assigned to "disagree" and "disagree strongly" responses, while only 21 per cent are assigned to "agree" and "agree strongly" responses. Hence, if acquiescence set is general over tests, acquiescent subjects should get low MTAI scores. Our two significant negative  $r$ s support this reasoning.

Table 1 further shows that both acquiescence tests correlate positively with the F scale. The  $r$  between the Plus scale and the F scale is .44; that between the Information-True score and the F scale is .35. These positive  $r$ s were expected by the same reasoning as that given for the negative  $r$ s between acquiescence and the MTAI. On the F scale, however, the scoring weights correlate *positively* (+1.00) with degree of agreement on the items. Our positive  $r$ s confirm those of Cohn and Bass.

The final  $r$  in Table 1 is that between the MTAI and the F scale. This  $r$  should be negative because both tests are affected by acquiescence tendency but score it in opposite directions, and because the content of the two tests is fairly similar. The obtained  $r$  was  $-.62$ .

#### PSYCHOLOGICAL SIGNIFICANCE OF ACQUIESCENCE

So far, our data merely support previous findings concerning the acquiescence set's

internal consistency, generality, and effect on test intercorrelations. At this point, however, we can choose between two interpretations of these findings. That is, granted that the California F scale and the MTAI are loaded with acquiescence set, we can say either (a) that acquiescence is psychologically irrelevant to the authoritarian attitude measured by the F scale and the MTAI, or (b) that acquiescence is part and parcel of the authoritarian attitude. The latter position, developed by H. J. Leavitt and others (13), is that which we shall adopt, extend, and further test in the remainder of this study.

Not only the MTAI and the F scale are involved in this issue. As Bass (4) has indicated, measures of many other attitudes are implicated. These measures involve attitudes about social relations, morals, prejudice, custom, and status relations; examples are "dogmatism," "misanthropy," "xenophobia," and "suggestibility."

#### Positive and Negative Items

*The act of agreeing.* Let us assume that acquiescence tendency reflects a genuine aspect of the authoritarian disposition. Then agreeing with a statement should, other things being equal, reflect authoritarianism. But suppose the statement agreed with is one that expresses a nonauthoritarian attitude, i.e., is a "positive" statement. Examples of positive statements can be found in Bass's G scale, and in Ancona's  $\frac{1}{F}$  Scale (2); both of these were made up of items opposite in meaning to the original statements of the F scale. Examples of these items are the following: "The artist and the professor are much more important to society than the businessman and the manufacturer." "Human nature being what it is, universal peace will come about eventually."

The MTAI contains 38 such "positive" items, where agreement may be considered to reflect a nonauthoritarian attitude. Examples from the MTAI are: "Minor disciplinary situations should sometimes be turned into jokes." "Unquestioning obedience in a child is not desirable." "There is too great an emphasis upon 'keeping order' in the classroom."

As shown in Table 1, the act of agreeing to positive items can occur either (a) "logically,"



TABLE 2  
HYPOTHESIZED EQUIVOCALITY OF RESPONSE TO POSITIVE  
ITEMS AND UNIVOCALITY OF RESPONSE TO  
NEGATIVE ITEMS

Type of Item	Response	
	Agree	Disagree
Positive	Logical Meaning: Non-authoritarian Psychological Meaning: Authoritarian	Logical Meaning: Authoritarian Psychological Meaning: Nonauthoritarian
Negative	Logical Meaning: Authoritarian Psychological Meaning: Authoritarian	Logical Meaning: Non-authoritarian Psychological Meaning: Nonauthoritarian

because the person genuinely has the non-authoritarian attitude which the content of the item represents, or (b) "psychologically," because he tends to acquiesce. Now, if acquiescence reflects obeisance to authority, i.e., a form of authoritarianism, then reasons *a* and *b* have opposite psychological significance. The "logical" and "psychological" reasons for agreeing are in a sense opposed to each other. Response to a nonauthoritarian statement may therefore be considered to have equivocal meaning for the personality of the subject involved.

Table 2 also shows that on negative items, however, agreeing reflects authoritarian attitudes unequivocally, i.e., both "logically" and "psychologically." Examples of negative items are all those of the F scale, such as: "People can be divided into two distinct classes, the weak and the strong." "Familiarity breeds contempt." "War and social troubles may someday be ended by earthquake or flood that will destroy the whole world."

There are 112 such negative items in the MTAI: "If the teacher laughs with the class in amusing classroom situations, the class tends to get out of control." "A child should be taught to obey an adult without question." "Discipline in the modern school is not as strict as it should be."

*The act of disagreeing.* On positive items, disagreeing may occur (c) "logically," because the person is authoritarian and so genuinely rejects the nonauthoritarian attitude which the content of the statement represents, or (d) "psychologically," because he is nonauthoritarian and so does not tend to acquiesce ir-

rationally when confronted with a difficult or ambiguous statement. Here the person does not have the kind of obeisance to authority of the printed word which may be considered to reflect an aspect of the authoritarian personality. The nonauthoritarian is not negativistic; he simply does not agree as often on merely temperamental grounds, and his alternative therefore is to disagree more often. For positive items, again, reasons *c* and *d* have opposite significance.

On negative items, however, the act of disagreeing has the same significance for authoritarianism whether it is a genuine response to item content or a temperamental response representing autonomy, nonyielding, and the like. In short, agreeing or disagreeing with negative items has relatively univocal meaning for the authoritarianism of the person being tested.

*Hypothesis.* By this reasoning, we arrive at the hypothesis that *negative items have greater validity than positive items for the measurement of authoritarianism.* Psychologically, disagreeing requires more "ego strength," "emotional security," and "self-confidence," than does the act of agreeing. Coupled with the logical significance of a nonauthoritarian response, i.e., of disagreeing with a negative statement, the meaning of the response is clearly nonauthoritarianism. Psychologically, agreeing may mean yielding to the authority of the printed word or a similar submission to the power and plausibility of the social context in which the test is given. Coupled with the logical significance of a nonauthoritarian response, i.e., of agreeing with a positive statement, the resultant meaning of the response is relatively unclear; it may mean either authoritarianism or nonauthoritarianism. When both positive and negative items are similar in content and degree of ambiguity, the latter items should have higher validity for measurement of authoritarianism.

### Method and Results

We have tested this hypothesis by correlating authoritarianism scores obtained separately with the positive and negative items of the MTAI, with ratings of teachers by their pupils. To test our hypothesis requires an external criterion of authoritarianism, unfounded with the subject's acquiescence. As such a measure, we used pupils' mean ratings of their teachers on Leeds' "My Teacher" rating scale. This scale consists of 50 Yes-No-? items of the following type:

"Is this teacher often 'bossy'? Does this teacher force her ideas on the pupils? Is this teacher usually fair with the pupils? Does this teacher think she is always right and the pupils always wrong?" In previous studies (6), pupils' ratings of teachers on this scale have been found highly internally consistent over items and highly discriminating between teachers. Further, these mean ratings by pupils have been significantly correlated with experts' and principals' ratings of teachers as to their ability to maintain discipline, create a friendly classroom atmosphere, establish a feeling of security on the part of pupils, exert a stabilizing influence on the class, and develop the pupils' self-reliance. It seems defensible to use these ratings as a criterion of the authoritarian personality trends of the type at which the F scale and the MTAI are aimed. The authors of the MTAI cite *The Authoritarian Personality* as an influence on their description of the adjustment mechanisms responsible for undesirable teacher attitudes.

The subjects in this study were 97 of the 98 teachers of grades 4-6 in a midwestern city (10). The criterion ratings of the teachers were furnished by their 2800 pupils. The reliability of the mean rating of teachers was .91, as estimated with Horst's formula (11).

Table 3 shows the correlations of the scores on the positive and negative MTAI items with the pupils' ratings of the teachers. The  $r$  of the 38 positive items was .13. The 112 negative items were divided, by means of a table of random numbers, into three sets of 38, 38, and 36. This procedure made the scores on negative items comparable as to length of test with the scores based on positive items. These three sets of negative items correlated .30, .22, and .22 with the pupils' ratings. The mean  $r$  was .25.

The  $t$  value of the difference between  $r$ s of .13 and .25, using Hotelling's formula for the  $t$  of a difference between  $r_{12}$  and  $r_{13}$  (12, p. 278), is 1.64. Since the direction of the difference was predicted, a one-tail level of significance is justifiable; hence, the difference is almost significant at the .05 level. This test of significance did not take account of the fact that the value of  $r$  of .25 is a mean of three  $r$ s and hence is more reliable than a single  $r$ .

When all 112 negative MTAI items were combined, the score on them correlated .31 with the pupils' ratings.

This  $r$  is higher than the  $r$  of .26 obtained with scores based on the entire 150-item MTAI. Thus, the score based only on negative items is more valid than the entire inventory.

## DISCUSSION

The main point of this paper is that acquiescence may belong to the family of dispositions—including authoritarianism, obedience to authority, and conformity—that have been identified in studies employing methods other than agree-disagree items. Accordingly, the agree-disagree format and high loading with negative items of the F scale and the MTAI make a positive contribution to the validity of these inventories.

On the basis of this reasoning, we hypothesized that, on positive items, the "logical" significance of the response might be contradicted by the "psychological" significance of the act of agreeing or disagreeing. Hence, such items should be less valid than negative items, in which the act of agreeing or disagreeing represents psychologically the same authoritarianism or nonauthoritarianism as the logical significance of the response denotes. Positive support for this reasoning was yielded by the correlations of the two types of items in the MTAI with pupils' ratings of their teacher on a variable akin to authoritarianism.

Our results resemble those described by Rundquist (15, pp. 142-143) and Cronbach (7, p. 406). Rundquist cited studies showing negative opinion items more valid in discriminating between men receiving and not receiving public relief, between employed and unemployed engineers, between problem and normal children. Cronbach found that number-right scores on False statements on True-False achievement tests were more valid than scores on True statements against criteria consisting of grades on other types of examinations.

These results can be assimilated to ours only if acquiescence is considered to reflect not only obedience and authoritarianism but also maladjustment, low morale, and low intelligence. Granted such assumptions, we could prepare diagrams like Table 2 to show that responses to positive items are equivocal in terms of these variables, while responses to negative items are not. That these assumptions are plausible is indicated by some of Crutchfield's findings: "The assessment staff rating on 'intellectual

TABLE 3  
CORRELATIONS OF TWO TYPES OF MTAI SCORE  
WITH PUPILS' RATINGS OF TEACHERS  
( $N = 97$ )

Type of MTAI Score	No. of Items	M	SD	$r_{tt}^{\dagger}$	$r$ with Mean Rating of Teacher by Pupils
Positive Items	38	10.3	11.8	.75	.13
Negative Items					
Set A	38	14.8	10.2	.82	.30
Set B	38	17.7	9.3	.75	.22
Set C	36	14.2	8.9	.73	.22
All items	112	46.8	26.0	.91	.31
Total Test	150			.90	.26

$\dagger$  Corrected split-half.



competence' correlates  $-.63$  with conformity score, this being the highest relationship of any found... An 'ego strength' scale... correlates  $-.33$  and a staff rating on 'leadership ability,'  $-.30$  with conformity" (9, p. 194). Similarly, Ancona's (2) finding that acquiescence is inversely related to need achievement would support our assumption.

Further tests of this hypothesis should be made. They should employ both kinds of items—positive and negative. Scores for authoritarianism on both types of items should be validated against external criteria of authoritarian personality tendencies in which the acquiescence set is not involved. Similarly, measures of acquiescence should be correlated against measures of morale, ego strength, conformity, and intelligence.

If negative items prove to be more valid for the measurement of authoritarian personality tendencies, this will have practical implications for test construction. Either inventories should consist primarily of negative items, as does the present MTAI; or most positive items should be unscored; or scores on positive and negative items should be obtained separately and weighted empirically.

It appears, with the hindsight provided by the present data, that the authors of the F scale and the MTAI built better than they knew. No mention of the acquiescence variable or of its possible kinship with authoritarianism is made in the descriptions of either scale's construction. Levinson (1, p. 59) in describing the construction of the anti-Semitism scale, gave as his reasons for the use of negative items (a) they tend to be more discriminating, (b) they can be so phrased that they express subtle hostility... (c) they are more reasonably used for measuring receptivity to anti-Semitic ideology, (d) evidence showed variability of response rather than a set to agree or disagree consistently, and (e) the scale was valid against external criteria. None of this reasoning considers the psychological difference between agreeing and disagreeing when confronted with a difficult or ambiguous item. Until now, the argument *against* using only negative items has not been that they will produce a uniform set, but rather that they will reflect *individual differences* in such a set. Previous writers have shown that this does in fact

occur. Our findings show that it not only occurs, but that it probably contributes positively and in a psychologically meaningful way to the validity of the F scale.

The present results also have theoretical significance for the nature of acquiescence, authoritarianism, and related variables. Our results provide some support for the conceptualization of acquiescence as akin to the "yielding" or "conformity" to group pressure studied by Asch (3) and Crutchfield (9). Crutchfield (9, p. 194) reported an  $r$  of  $.39$  between conformity and the F scale. Rudin (14) found the F scale correlated about  $-.35$  with freedom from contextual influence in perception of both social and nonsocial stimuli. If acquiescence does represent a kind of conformity—in this case to the "pressure" of the printed word or of the context in which a plausible statement is presented—then it does represent an aspect of authoritarianism.

In short, the present study may aid in understanding the psychological meaning of acquiescence set, apart from its logical and mathematical significance. Acquiescence has long been suspected of being a psychologically meaningful variable. It appears that, in measuring authoritarianism, we do have one of the potentialities to which Cronbach referred: "Some of the response-set variance is potentially useful, some of it is an interference with measurement. The problem for the tester is to capitalize on the effect of response sets where they are helpful to validity..." (8, p. 19).

Further studies are needed to pin down the proposition that acquiescence set should be capitalized on, rather than shunned, in tests of authoritarianism, ethnocentrism, teacher attitudes, and the like.

## SUMMARY

Previous workers have hypothesized and found significant correlations between the California F scale and acquiescence set (reliable individual differences in tendency to respond "agree" or "true" when in doubt). To corroborate these findings and extend them to the Minnesota Teacher Attitude Inventory (MTAI), we constructed an information test of acquiescence, consisting of 50 true-false items about obscure matters. This test, Cohn's Plus Scale of acquiescence, the F scale, and the

MTAI all correlated significantly with each other in directions expected on the basis of the acquiescence set.

The question was then raised, is acquiescence an aspect of authoritarianism such that variance due to acquiescence set is psychologically relevant to authoritarianism? By assuming that acquiescing reflects authoritarian personality tendencies, we can derive the hypothesis that "positive" items are less valid than "negative" items for measuring authoritarianism. Teachers' scores on these two types of MTAI items correlated differently, as hypothesized, with pupils' ratings of the teachers on variables similar to authoritarianism.

The hypothesis implies that the F scale, the MTAI, and similar inventories are advisedly constructed primarily of "negative" items. The psychological meaning of acquiescence resembles that of authoritarianism, conformity, low ego strength, and low intelligence.

#### REFERENCES

1. ADORNO, T. W. *et al. The authoritarian personality*. New York: Harper, 1950.
2. ANCONA, L. Indagine sulla natura psichica del "Response Set." La "motivazione al successo." *Arch. psicol. neurol. Psichiat.*, 1954, **15**, 23-74.
3. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
4. BASS, B. H. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.

5. COHN, T. S. The relation of the F scale to a response to answer positively. *Amer. Psychol.*, 1953, **8**, 335. (Abstract)
6. COOK, W. W., LEEDS, C. H., & CALLIS, R. *Minnesota Teacher Attitude Inventory: Manual*. New York: Psychological Corp., 1951.
7. CRONBACH, L. J. Studies of acquiescence as a factor in the true-false test. *J. educ. Psychol.*, 1933, **401-415**.
8. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3-31.
9. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191-198.
10. DELLA PIANA, G. M., & GAGE, N. L. Pupils' values and the validity of the Minnesota Teacher Attitude Inventory. *J. educ. Psychol.*, 1955, **167-178**.
11. HORST, P. A generalized expression for the reliability of measures. *Psychometrika*, 1949, **21-31**.
12. HOTELLING, H. The selection of variates for use in prediction with some comments on the problem of nuisance parameters. *Ann. math. Stat.*, 1940, **11**, 271-283.
13. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with the authoritative. *J. Psychol.*, 1955, **40**, 215-221.
14. RUDIN, S. A. The influence of context on the perception of complex stimuli. Unpublished doctoral dissertation, Univ. of Illinois, 1955.
15. RUNDQUIST, E. A. Form of statement in personality measurement. *J. educ. Psychol.*, 1939, **31**, 135-147.

Received September 24, 1956.



# CULTURAL INFLUENCE ON THE PERCEPTION OF MOVEMENT: THE TRAPEZOIDAL ILLUSION AMONG ZULUS<sup>1</sup>

GORDON W. ALLPORT AND THOMAS F. PETTIGREW

Harvard University

**T**RADITIONALLY, theories of the visual perception of movement—with which the present study deals—have been divided into two classes: 1. The *nativistic*, i.e., theories emphasizing the role of retinal and cortical functions relatively unaffected by learning, habit, experience, or meaning; and 2. The *empiricistic*, i.e., theories giving primary weight to the role of experience and learning.

For our purposes it is essential to subdivide empiricistic theories into two groups:

*Cumulative habit:* Stressing the effects of many types of early, remote, and generalized experience which by transfer or cross conditioning become a major determinant of the perception of movement. Toch and Ittelson (22, p. 199) state that "contemporary empiricism" favors this type of approach, offering its explanations of perceived movement in terms of "weighted averages of experiential sediments of all kinds acting inseparably."

*Object connotation (meaning):* Explaining perceived movement largely in terms of familiar objects. One sees continuous wing motion in an electric sign representing a bird in flight, although the stimulus actually occurs discontinuously in two or in three fixed positions. This theory would hold that our familiarity with birds in flight causes us to fill the gaps with perceived motion. A good statement of this theory of stroboscopic movement

may be found in James (9). This author insisted that "perception is of definite and probable things." In explaining illusions, James leaned heavily upon their resemblance to familiar objects. In so doing he was merely rendering more concrete and specific Helmholtz's theory of "unconscious inferences" and Wundt's "assimilation" theory.

At the present time none of these theories can be defended in their exclusive purity. F. H. Allport (1) makes this fact abundantly clear. No empiricist, for example, can deny the native physiological substrate of the perception of movement nor its structural properties as represented by Korte's laws governing the phi-movement (13). Conversely, no nativist can deny the contribution that past experience may make to perceived movement. Wertheimer, for instance, admits that "it is certainly correct that past experience can influence the conditions (*Verhältnisse*) of stroboscopic vision" (23, p. 79). Von Schiller (21, p. 195) makes the suggestion—especially important for the present research—that attitudinal set and expectancy are particularly effective in determining the perception of movement in ambiguous (*alternativ*) situations.

Many authors who have worked with perceived movement, e.g., Neff (17) and Hall and Earle (7), have favored an eclectic view. They have concluded that native and experiential factors both contribute, as do likewise momentary set and the previous level of adaptation. And they allow, among experiential factors, for both *cumulative habit* and *object connotation* (meaning).

While the eclectic position is no doubt correct there is, as Toch and Ittelson insist (22), still a fundamental question concerning "the relative primacy or importance of learning processes and physiological conditions." Theorists do tend to lean in one direction or the other. The distinction between nativists and empiricists still exists. Among the former, one thinks of Wertheimer (23), Koffka (12), Metzger (14), Michotte (15); among the latter, of Helmholtz (8), Piéron (19), Ames (2), Wundt (24), Cantril (4), G. E. Müller (16),

<sup>1</sup> For their services as interpreters we wish to thank Kate Mdlala, Ethan Mayisela, Emmanuel Zuma, and Mr. September. For assistance in conducting the experiments we are indebted to Eric Barnett, Peter Hey and Ann Pettigrew. In making arrangements for the investigations we received kind cooperation from the Native Affairs Commissioner of the Province of Natal, from Dr. John Bennett of the Polela Health Center, Dr. Wolfgang Bodenstein of the Ceza Medical Mission, Max Marwick of the Institute for Social Research of the University of Natal, Dr. B. Gampel of the Natal Medical School, and Mr. J. Mngadi of the Lamontville Community Center; also from the Paramount Chieftainess, Thoko C. Zulu. For the opportunity to spend several months as members of the Institute for Social Research at the University of Natal, we are grateful to the Principal, Dr. E. G. Malherbe.

The original suggestion for the experiment grew out of conferences with Hadley Cantril for whose strategic advice we are grateful. We benefited likewise from discussion with F. P. Kilpatrick and E. Engel. The investigation received financial support from a grant from the Rockefeller Foundation to Princeton University for research in perception.

Kilpatrick (11), and Toch and Ittelson themselves (22).

### THE CROSS-CULTURAL APPROACH

To gain light on this dispute psychologists have often asked, "How about primitive peoples?" If we can find a tribe or a culture where relevant past experience can be ruled out, we could then determine whether the perception resembles that of western peoples. If it does so, then the argument for nativism is presumably stronger. The first extensive attempt to apply this test was made by W. H. R. Rivers during the Torres Strait expedition in 1898 (20). Rivers presented to the island natives a whole array of visual illusions and compared their reports with western norms. For some of the illusions there were no appreciable differences; for others, the natives seemed on the whole less susceptible than westerners. While Rivers himself does not make the point clearly, his results seem to show that illusions involving object connotation (e.g., a European street scene) are far less compelling to the natives than are illusions having no such object connotation (e.g., the rotating spiral) (20, esp. pp. 130 f.).

It is not easy for western psychologists to visit primitive tribes, nor to conduct among them adequately controlled experiments. The present article, however, deals with one such attempt. But before we describe it, the theoretical point at issue should be made entirely clear: *We do not claim to be testing the merits of the nativist or empiricist positions directly.* For reasons that will appear in the course of our discussion, we do not believe that comparative perceptual studies on western and on primitive peoples can solve this particular riddle. *We claim only to have illuminated the part played by object connotation (meaning) in the perception of motion as over and against the part played by either nativistic determinants or cumulative habit.* Our experiment is not able to distinguish between the role of these last two factors.

### The Rotating Trapezoidal Window

Before the days of Gestalt psychology it was customary to regard visual illusions as oddities, as exceptional experiences to be accounted for either in terms of nativistic or experiential constraints. Today, however, we make little distinction between illusions and

veridical perceptions, since no illusion lacks veridical elements and no veridical perception is devoid of subjective shaping. So-called illusions are simply instances of perception where the discrepancy between impression and knowledge (whether the knowledge be the subject's or the experimenter's) is relatively striking. It is in such "looser" conditions of perception that theorists often seek to obtain light on the relative weight of factors entering into the normal perceptual process. The reasoning is not unlike that which leads psychologists to study exaggerated functions in psychopathology in order to obtain light on the same but less exaggerated functions of the normal mind.

Our experiment follows this logic, making use of the rotating trapezoidal window described by Ames (2)—a device that has been called "a dramatic masterpiece of ambiguous stimulation" (1, p. 276). The window (Figure 1) is so proportioned that as it rotates, the length of the longer edge is always longer on the retina than is the shorter edge (even when the shorter edge is nearer). The resulting perception is normally one of oscillation or sway; the observer apparently tending to keep the longer edge nearer to him. Instead of seeming to rotate, as it actually does, the window is seen to sway back and forth in an arc of 90 to 180 degrees.

An appended cube and rod add great interest to the illusion, since the perceived *rotating* of these objects conflicts sharply with the perceived *sway* of the window. In consequence, the cube is usually seen to detach itself and swing without support in a ghostly fashion in front of the window (for that period of time when the shorter edge, to which it is attached, is in fact nearer to the subject). Similarly, the rod bends, twists or "cuts through" the mullions in order to accommodate itself to the phenomenal oscillation. The observer finds the bizarre effect both amusing and inexplicable.

The window used in the present experiment is smaller than the original Ames window; length  $13\frac{1}{4}$  inches, height of the long side  $12\frac{1}{2}$  inches, height of the short side 9 inches. Ames demonstrated that within limits the ratio of the sides of the trapezoid to one another cannot affect the illusion. The optimum speed of rotation Ames reports as 3 to 6 r.p.m. Our own motor driven window ran slightly less



than 2 r.p.m. The original Ames window had mullions dividing it into 15 frames; ours had 6 frames (probably more normal for a "window"). For certain comparisons, we employed also a true rectangular window, 12" x 10½".

The explanation Ames gives for the illusion maintains (a) that the observer, owing to familiarity with rectangular windows assumes *this* window to be rectangular; and (b) that owing to long experience with doors, windows, and similar objects, the observer has learned to interpret longer retinal stimulations as coming from nearer objects. Hence, the longer edge of the window is interpreted as being nearer, and the window is seen to oscillate rather than to rotate.

Ames (2, p. 14) gives a clearly empiricistic explanation with a leaning toward the object connotation version:

In his past experience the observer, in carrying out his purposes, has on innumerable occasions had to take into account and act with respect to rectangular forms, e.g., going through doors, locating windows, etc. On almost all such occasions, except in the rare case when his line of sight was normal to the door or window, the image of the rectangular configuration formed on his retina was trapezoidal. He learned to interpret the particularly characteristic retinal images that exist when he looks at doors, windows, etc., as rectangular forms. Moreover, he learned to interpret the particular degree of trapezoidal distortion of his retinal images in terms of the positioning of the rectangular form to his particular viewing point. These interpretations do not occur at the conscious level, rather, they are unconscious and may be characterized as *assumptions* as to the probable significance of indications received from the environment.

It should be added that Ames does not insist that object connotation ("windowness") is the sole determinant of the illusion. He himself employed a variety of trapezoidal figures and discovered that even a plane surface of trapezoidal shape arouses the illusion of sway, though to a much less degree than does a "window frame" (2, p. 29).

### *The Hypothesis*

In order to test the "object connotation" theory, we studied various groups of Zulu children (10-14 years old) in Natal whose own culture is virtually devoid not only of windows, but, to a surprising extent, of angles, straight lines, and other experiential cues that would presumably "cause" the illusion if it were wholly a product of experience. Our hypothesis therefore is:

*Zulu children, provided they are unacculturated (amabinca) will report the illusion of sway in the trapezoidal window less often than will urbanized acculturated Zulu children (amabunguka) or than white ("European") children.*

### *The Zulu Culture*

Zulu culture is probably the most spherical or circular of all Bantu cultures, possibly the most spherical of all native African cultures (though it would be difficult to prove this contention). The word "zulu" means heavens or firmament, and the aesthetic ideal of round rather than angular styles affects native art, architecture, and speech.

Huts are invariably round (rondavels) or else beehive shaped, whereas in other Bantu tribes they are sometimes square or rectangular. Round huts arranged in a circular form with round stockades to fence in animals, constitutes a typical African homestead (kraal). Fields follow the irregular contours of the rolling land, and never seem to be laid out in the neat rectangular plots so characteristic of western culture. See Figure 2.

The typical Zulu hut has no windows, and no word for such an aperture exists. In the more primitive beehive grass huts, doors are merely round entrance holes; in the round mud huts doors are amorphous, seldom if ever neatly rectangular. Cooking pots are round or gourd shaped. In his studies among Zulus, L. Doob (6) finds that the less acculturated natives, relative to westernized natives, show a statistically significant preference for circles over squares when they are asked to choose between designs drawn in these shapes (personal communication to the authors).

It is commonly said in Natal that Zulus fresh from reserves cannot plow a straight furrow and are unable to lay out a rectangular flower bed. Such inability is of course overcome with experience and training, but the initial defect would seem clearly related to the circularity that is characteristic of life on the reserves and to the lack of familiarity with straight layouts.

Linguistically, the same bias towards circularity is seen. While it is possible to say "round" in Zulu, there is no word for "square." There is a word for "circle" but not for "rectangle." To speak of window, of square, or of



FIG. 1. THE EXPERIMENTAL SITUATION



FIG. 2. CIRCULARITY IN THE ZULU ENVIRONMENT



rectangle at all, a Zulu is forced to borrow these terms from Afrikaans or from English—provided he is able to do so.

### *The Subjects*

The experiment required the use of two contrasting groups of subjects (*Ss*): those who had lived all or most of their lives in western culture, and those who were unacculturated. Even in the Bantu reserves or in Zululand itself it is not possible to make certain that a resident does not know what a window is like. While schools, churches, and health centers are few and far between, they are nevertheless within the possible range of visitation by most native inhabitants, even children. Our experiments at Polela and Ceza took place in health centers, at Nongoma in a court house. The *Ss*, to be sure, were brought in from remote parts of the reserves by lorry, or came on foot; but they had at least this one-time acquaintance with a rectangular building and windows.

Still, it is possible to say that the experiment dealt with two widely contrasting groups in respect to the degree of experience they had had with western architecture and ways of life. Some members of the more primitive groups, for example, may never have seen windows with rectangular panes of glass prior to the actual experimental situation.

By using herd boys as *Ss*—mostly between 10 and 14 years of age (few of them knew their age exactly)—we were able to make certain that they had never been off the reserves and had never attended school. Boys of the same age comprised our urban control groups: one group of European boys at Greyville Community Center; another group of Bantu boys at the Lamontville Community Center in Durban. Most of these urban boys were attending school.

Our major experiment thus involved the following groups:

- |         |                        |            |
|---------|------------------------|------------|
| Group A | Urban European boys    | (20 cases) |
| Group B | Urban African boys     | (20 cases) |
| Group C | Polela Rural Africans  | (20 cases) |
| Group D | Nongoma Rural Africans | (20 cases) |

A rough indication of the cultural differences between the rural and urban groups lies in answers to the question asked at the end of the experiment about the rectangular window, "What does this look like?" The percentage saying "window" or "window frame" among the urban children was 88; among the rural, 45.

### *Procedure*

The procedure involves four conditions, varying two factors bearing on the perception of the illusion: monocular vs. binocular viewing and distance from the stimulus object. Each *S* saw first the rectangular, and then the trapezoidal window in at least 3 full revolutions under each of the following conditions.

- |               |        |           |
|---------------|--------|-----------|
| First trial:  | 10 ft. | binocular |
| Second trial: | 10 ft. | monocular |
| Third trial:  | 20 ft. | binocular |
| Fourth trial: | 20 ft. | monocular |

It was thought that this order would impose the "hardest" condition first and therefore minimize the effects of suggestion. One might fear that if at 20 feet

with one eye a subject easily perceived the illusion he might become accustomed to expecting oscillation in the trapezoidal window at closer distances and under binocular conditions. Conversely, of course, it might be argued that a subject who cannot perceive the illusion at ten feet binocularly would form an expectation that might prevent his obtaining it under easier conditions. We shall refer later to a control experiment (starting at 20 feet monocularly) designed to check on any suggestive effect that might arise from our order of presentation.

The experimenter (*E*) required the assistance of a second psychologist who kept records of the *Ss*' reports, also of an interpreter with all African *Ss*. Care was taken to prevent *Ss* who had finished the experiment from communicating with *Ss* who had not.

After being put at ease, the *S* gave his age (if he knew it) and his degree of education (if any). The *S* then sat in a chair placed at the proper distance from the window and was told to watch carefully the movement that he would see. After approximately three revolutions the *E* asked, "How does it seem to you to be moving?" Often the *S* spontaneously used his hands to indicate the motion until the *E* was satisfied whether a full rotation or a fluctuation was intended. The use of hand motion by the *S* proved to be fully convincing, for when he reversed the hand at precisely the right moment for the illusion to occur there could be no question concerning his experience. This device gave a useful check on the accuracy of the translator's report of the *S*'s verbal statements.

After obtaining a report for the rectangular window in each of the positions, the trapezoidal window was inserted in place of the rectangular, and the same method of report employed. In addition, the *S* was asked to tell whether the motion of the trapezoidal window was "like" that of the first window. This procedure served as a further check on the verbal description and hand report. In nearly all cases it was possible to record a clear and unequivocal judgment of the *S*'s perception. Less than three per cent of all judgments were listed by the *E* as "uncertain."

Whenever the illusion was reported for the first condition, the bar was inserted and the *S* asked, "How does the bar move?" and "Does the bar stay straight?" On occasional trials when the *S* had reported both the sway of the window and the bending of the bar, the cube was attached and the *S* asked to describe its motion. In these cases there was usually laughter (as with American *Ss*) and considerable confusion and difficulty manifested in describing so unreal and "spooky" a motion. Because of the difficulty of communicating concerning these complex phenomena we make no further systematic use of the cube and rod in the present study.

At the conclusion of the experiment, the *S* was asked what the rectangular window "looked like." He also stated his preference for one of two geometrical drawings presented to him in pairs (a circle, square, trapezoid). He then received a slight payment for his services (usually one pound of sugar or a candy bar and sixpence).

### RESULTS

*General results.* Table 1 gives the results for the two unacculturated groups (Nongoma and

TABLE 1  
NUMBER REPORTING ILLUSION  
(Boys 10-14 Yrs. of Age,  $N = 20$  in Each Group)

Condition	Nongoma Rural			Polela Rural			African Urban			European Urban		
	Yes	No	Uncertain	Yes	No	Uncertain	Yes	No	Uncertain	Yes	No	Uncertain
First condition (10', both eyes)	3	17	0	4	14	2	13	7	0	11	9	0
Second condition (10', one eye)	14	6	0	16	4	0	19	1	0	19	1	0
Third condition (20', both eyes)	8	12	0	17	1	2	16	3	1	16	4	0
Fourth condition (20', one eye)	18	2	0	17	2	1	18	2	0	19	0	1
Totals	43	37	0	54	21	5	66	13	1	65	14	1

Polela Reserves) and for the two districts within metropolitan Durban, African (Lamontville), and European (Greyville).

Combining all four conditions, there is a very significant tendency for the urban groups to report the illusion more often than the rural groups (corrected  $2 \times 2 X^2 = 15.34$ ;  $p < .001$ ). This difference is most marked with the first condition (corrected  $2 \times 2 X^2 = 12.38$ ;  $p < .001$ ). There is also a significant trend with the second condition (corrected  $2 \times 2 X^2 = 4.80$ ;  $p < .05$ ) and a slight tendency with the third condition (corrected  $2 \times 2 X^2 = 1.87$ ;  $p < .20$ ) for the rural children to observe the illusion less often than the urban children. Virtually no difference exists with the fourth, 20 feet and one eye condition.

Table 2 expresses the results in an alternative way. Since four conditions of presentation were used we can determine in how many of these four conditions on the average each of the cultural groups reported the illusion. For the two unacculturated groups combined, the illusion is reported in 2.425 of the four conditions, while for the acculturated groups the average is 3.275. This mean difference has high statistical significance ( $t = 3.51$ ;  $p < .001$ ).

It is evident from Tables 1 and 2 that city dwellers, whether Zulu or European, find the illusion somewhat more compelling than do rural ("primitive") natives. This tendency is especially pronounced at 10 feet with binocular vision—a condition when binocular cues of true depth (in this case, true rotation) are most plentiful. The reader will also note that the results for Polela (rural) stand somewhat between those for the city children and those from Nongoma (rural). At ten feet binocularly, they resemble those of Nongoma; at twenty feet binocularly, those of the city boys. Polela

TABLE 2  
DISTRIBUTION OF SCORES

Sample	Number of Yes's					Average
	4	3	2	1	0	
Nongoma	2	4	10	3	1	2.15
Polela	4	10	3	2	1	2.70
Urban African	12	4	2	2	0	3.30
Urban European	11	5	3	0	1	3.25
Total ( $N = 80$ )	29	23	18	7	3	2.85

is, in fact, one hundred miles closer to Durban than is Nongoma which lies in the heart of Zululand. There is no doubt that the children in Polela have somewhat more familiarity with western architecture (specifically with windows) than do the children of Nongoma. The results (Table 2) correspond to a continuum of cultures: city children having a maximum of familiarity with western architecture, Nongoma children the least.

*Preference for circles.* Following the experiment, all Ss were shown drawings of a square, a trapezoid, and a circle (in pairs), and asked to express a preference. Table 3 indicates that those who expressed a preference for the circle (at least once in the two pairings) tend in the African groups to report the illusion *less* often. This tendency holds for all experimental conditions for all three African groups. The relationship is statistically significant, however, for only the binocular conditions. Circle-preferring Zulu children report the illusion significantly less often than the angle-preferring Zulus in Conditions 1 and 3 (corrected  $2 \times 2 X^2 = 3.89$ ;  $p < .05$ ), but the difference in the monocular, second and fourth, conditions is not significant (corrected  $2 \times 2 X^2 = 0.18$ , *n.s.*). There are no differences ap-



TABLE 3

PERCENTAGE OF CASES REPORTING ILLUSION AMONG SUBJECTS PREFERRING AND NOT PREFERRING CIRCLE

Condition	Combined African Groups N = 60		European Group N = 20	
	Preferring Circle N = 39	Not Preferring Circle N = 21	Preferring Circle N = 12	Not Preferring Circle N = 8
10' binocular	28	43	58	50
10' monocular	79	86	100	88
20' binocular	59	86	83	75
20' monocular	87	90	100	88
All conditions	63	76	85	75

proaching significance between the circle and noncircle-preferring European Ss.

Let us assume that the aesthetic preference for circles may provide an index of the subjective closeness of the individual to Zulu culture (since it is, as we have seen, overwhelmingly a circular culture). If we do so we may say that this subjective closeness seems to predispose the *S* to resist the illusion. Stated in terms of transactional theory, rectangles and trapezoids have less functional significance for him. His preception of the window's rotation is accordingly more frequently veridical.

We have noted that this influence is significant only in the conditions involving *binocular* perception. A reasonable interpretation would be that cultural effects cannot easily change the basic demand character of the illusion monocularly perceived, but may do so when binocular conditions leave more latitude for choice and for interpretation among a greater number of cues.

This result then, so far as it goes, lends some weight to the contention that "cultural significance" is playing an appreciable part in determining the results.

*Illusion with rectangle.* Before viewing the trapezoidal window, every *S* in all four conditions first saw the rectangular window rotating. The purpose was to make sure that the sway (oscillation) reported for the trapezoid was judged to be *different* from the motion of the rectangular window. In most cases, indeed, the *S* was able to make the distinction clearly, indicating by gesture and by words that the rectangular window went "round and round" whereas the trapezoid oscillated.

There were cases, however, where the rec-

tangular window was reported as oscillating. In fact, nearly one-third of the 80 Ss reported such a phenomenon at one or more of the four conditions. The actual percentage reporting sway in the *rectangular* window at each of the four conditions is:

First condition	0
Second condition	8
Third condition	16
Fourth condition	28

It is conceivable that this curious and somewhat unwelcomed finding may be a result of a "suggestive" order of presentation. Thus, no *S* seeing the rectangle before the trapezoid under the first condition (10 feet binocularly) reported the phenomenon. And, with the exception of 3 cases, no *S* reported the rectangular illusion in the second, third, or fourth condition *unless* he had previously reported the trapezoidal illusion. Altogether, 81 per cent of our Ss reported the illusion monocularly at 20 feet for the trapezoidal window, but only 28 per cent did so for the rectangular window under the same condition. In virtually all these cases the Ss had grown accustomed to seeing oscillation at some previous stage with the trapezoid.

Pastore (18), however, finds that more than half of his 58 American college Ss reported sway with the rectangle during a three-minute exposure, and at considerable distance from the window (where the retinal angle subtended by the two shapes may be subliminal). He does not tell whether the Ss had grown accustomed to the sway of the trapezoid before they reported sway in the rectangle. We must leave this problem for the time being unsolved.

#### A CONTROL EXPERIMENT

In order to determine whether unwanted suggestive effects, caused by our order of presentation, were influencing the results at the optimal stage for the trapezoidal illusion (viz., 20' monocularly) we simplified our procedure with entirely new Ss. Urban and rural Africans served as before. To secure the latter, we visited the Ceza Medical Mission in Zululand, approximately 20 miles from Nongoma. Both Ceza and Nongoma are in the deepest part of the native reserves, over 200 miles north of Durban. Polela, as we have said, lies about 100 miles west of Durban and has more European influence (e.g., European-

style architecture). This fact, we repeat, seems to explain why, as Tables 1 and 2 show, the Polela Ss report the illusion somewhat more frequently than do Ss at Nongoma or Ceza.

In the control experiment, the *S* was asked to cover one eye. Sitting at twenty feet from the rotating trapezoidal window he then described its motion (both in words and by hand motion). Later he was seated at ten feet from the object and using both eyes described the motion, comparing it with the previous motion. Finally he was, as in the other groups, asked his preference for the circular, square, or trapezoidal figures. None of these herd boys had ever been to school.

For an urban control group we used a fresh population of Lamontville boys of the same age range.

### Results

Table 4 gives the results.

These data are practically identical with those of the Nongoma and urban African samples cited previously. Again, the rural Zulu group reported the illusion significantly less often than the urban Zulu group in 10' binocular condition (corrected  $2 \times 2 \chi^2 = 11.53$ ;  $p < .001$ ), but no differences appear at the 20' monocular condition. This similarity of data proves that the order of presentation is not an important variable.

At the Ceza Mission Hospital we tested also a group of eleven expectant mothers, only one of whom had ever left the reserve. Eight reported the illusion at 20' monocularly, two did not, and one was uncertain. None of the eleven, however, reported it at 10' binocularly. These cases confirm the trend in all our tables that "primitives" are less able to perceive the sway in the trapezoidal window illusion under

TABLE 4  
NUMBER REPORTING ILLUSION IN CONTROL  
EXPERIMENT

Condition	Ceza (Rural) N = 24			Lamontville (Urban) N = 21		
	Yes	No	Uncertain	Yes	No	Uncertain
First condition (20', monocular)	22	2	0	20	0	1
Second condition (10', binocular)	2	18	4	14	7	0

TABLE 5

PERCENTAGE OF CASES REPORTING ILLUSION AMONG  
SUBJECTS PREFERRING TWO CIRCLES  
Combined African Groups (Ceza and Lamontville);  
N = 44

Condition	Preferring Two Circles N = 15	Preferring Less than Two N = 29
20' monocular	92	97
10' binocular	27	38

marginal conditions (i.e., at 10' binocularly) than are city dwellers.

Something should be said concerning the qualitative differences reported by Ss who first reported the illusion at 20' monocularly and then again at 10' binocularly. Not infrequently their reports at 10' binocularly were "mixed," that is to say, sometimes they saw the oscillation and sometimes not. In every case, the *S* was asked to tell the "difference" if any existed between the movement seen at 20' monocularly and that at 10' binocularly. Most Ss claimed that there was a difference: sometimes the window at the closer distance seemed to move faster, sometimes in a bigger arc, sometimes even in the reverse direction. And often, as we have said, the reports at 10' binocularly were "mixed—the subjects reporting sometimes a full rotation and sometimes a sway. We record "yes" to the illusion at 10' binocularly if at any point in the experiment the *S* reports a clear oscillation. Since the same criteria were applied at both Ceza and Lamontville, no source of error is introduced.

If the reader is acquainted with the illusion he will no doubt recognize this ambiguity in the perception at 10' when binocular cues are powerful evidence for true rotation, and yet the tendency to see illusory sway likewise exists. Because of this dual tendency we consider 10' binocularly as a *marginal* condition for the illusion. What is important for our purposes is the finding that under such marginal conditions urban children, who are familiar with western architecture, report the illusion much more frequently than do herd boys on the Zulu reserves.

As in the major experiment our Ss expressed their preference, in three paired comparisons, for a circle, trapezoid, or square. Since only seven cases of the 45 failed to choose the circle at least once, we changed our criterion



TABLE 6

PERCENTAGE OF SUBJECTS RECOGNIZING STIMULUS AS A "WINDOW" WHO REPORT THE ILLUSION

Condition	Recognizing "Window" N = 20	Not Recognizing "Window" N = 24
20' monocular	100	88
10' binocular	55	29

slightly from that used in Table 3. We determined the occurrence of the illusion among those who chose the circle *twice* as compared with those who chose it only once or not at all. Table 5 shows that in this population likewise, Ss who show a preference for circles tend—particularly in the binocular condition—to report the illusion somewhat less frequently than those who do not. The differences are not statistically significant but are in the same direction as those reported in Table 3. The implication of this finding, we repeat, seems to be that Ss whose aesthetic preference lies with the circularity of their tribal culture are the more resistant to the illusion.

At the conclusion of the experiment the investigator showed each S the rectangular window and asked, "What is this?" (In order to make certain that the children would have an opportunity to say "window" if they perceived the resemblance, the question was asked in three different ways in the Zulu language.) To one or more of these three questionings 67 per cent of the urban children said "window," but only 26 per cent of the Ceza children gave the same reply. If we combine all cases who said "window" at both Ceza and Lamontville we find an appreciable, though not statistically significant, tendency for them to report the illusion more often than do children who did not recognize the window-ness of the stimulus object (corrected  $2 \times 2 \chi^2 = 2.81$ ;  $p < .10$  when both conditions are combined). So far as it goes, this finding (Table 6) lends support to the object-connotation theory of the perception of movement, especially under the 10' binocular condition.

#### DISCUSSION

Our most striking finding is that under optimal conditions (monocularly at 20 feet) virtually as many primitive Zulus report the trapezoidal illusion as do urban Zulus or Europeans. Taking this one partial result by

itself we can say that the experiment supports either the nativistic or the cumulative habit theory. It does not by itself give us grounds for choosing between them.

Nativists might argue, for example, that whenever a longer and a shorter projection on the retina occur simultaneously the longer will assume a figure character and therewith a frontal position in the perception (other conditions being equal). Thus, some form of isomorphism obtains between retinal-cortical processes and the perception itself.

An empiricist with a "cumulative habit" preference might say that myriad ocular-motor adjustments from infancy have built up a dependable expectancy that longer projections on the retina will betoken nearer objects. One learns through repeated experience that longer retinal images of trees, cattle, people, stand for *nearer* objects (provided, of course, that one assumes such objects to be of equal size whether far or near from the eyes). It is not necessary for the S to have acquaintance with specific objects (in this case a window) in order to make a similar inference. The transfer effect is wide. Even the shadows painted on the rotating window are reminiscent of the S's experience with shadows in nature. Old experiences automatically condition novel experiences even though the latter are only analogous.

One assumption that may play a decisive part in this case is the assumption of "right angularity." From earliest life the child is conditioned to the fact that perpendicular objects best withstand the force of gravity. Circular though his culture is, his basic frame of reference is still one of verticals and horizontals. Seeing an entirely new object (the trapezoidal window) he assumes unconsciously (no less than do people who are familiar with windows) that its shape is rectangular. Just like people in western culture he may make this assumption even if he "knows" that the object is not in reality rectangular. This assumption, together with the assumption that longer objects on the retina are usually nearer objects, would predispose him to perceive that the longer edge of the window is always nearer (thus inducing the perceived oscillation). No less than people in western culture he would fail to "correct" his assumptions of right-angularity and of long-edges-

being-near-edges by his "knowledge" of the trapezoidal shape of the stimulus.

Our major result is clearly not compatible with a narrowly conceived object-connotation theory. It is not necessary for the *S* consciously to assume that the object is a window in order to experience the illusion. True, as Table 6 suggests, the specific object connotation seems somewhat to favor the illusion, but it is clearly not the decisive determinant. Thus, for example, 88 per cent of those who did not consciously recognize the frame as a "window" nevertheless experienced the illusion at 20' monocularly.

Yet, at the same time, our results show that object connotation cannot be disregarded. It also plays a part. Let us review the evidence:

1. Under all *suboptimal* conditions, as we see in Tables 1 and 4 (10 feet monocularly, and binocularly at 10 or 20 feet) there is a tendency for unacculturated *Ss* to report the illusion less frequently than do the acculturated.

2. The *Ss* who recognize the "windowness" of the stimulus object tend to report the illusion somewhat more frequently especially at 10 feet binocularly (Table 6).

3. African *Ss* expressing a preference for circles (assumed here to indicate a subjective closeness to the rotund Zulu culture) tend to report the illusion less often than those expressing preference for angular figures (Tables 3 and 5).

*We conclude that experience with, and identification with, western culture make it more likely that the illusion will be perceived under marginal (suboptimal) conditions.*

One fact reported by Ames, and mentioned above, supports our interpretation. He finds that a plane trapezoidal frame yields appreciable oscillation, but that the addition of mullions, panes, and shadows enhances the illusion. In other words, specific "thingness" contributes to the experience though it does not account for it wholly (6, pp. 28-31). And we may again allude to von Schiller's contention that expectancy is effective in determining perceived movement under marginal (*alter-nativ*) conditions.

May Brenner (3) likewise makes the point that when marginal conditions obtain, the *S* is forced to depend on stimulus *meaning*. On the other hand, when optimal stimulus conditions obtain, even brain-damaged cases report apparent movement to much the same degree as do normal cases.

Several other experiments have dealt with the effects of meaning on apparent movement. Thus, Jones and Bruner (10) report that in a stroboscopic experiment the line drawing of a man is seen to be in motion more actively than is a nonsense figure. De Silva (5) had previously established this same fact. Jones and Bruner conclude that "the more probable and practiced the movement, the more adequately will the movement experience maintain itself under suboptimal conditions" (10, p. 165). This conclusion is in agreement with our own.

Toch and Ittelson (22) report an experiment in which drawings of a bomb stroboscopically presented are seen in a downward (falling) motion, whereas drawings of an airplane presented in an identical fashion are seen in an upward (rising) motion. Though this experiment taken by itself favors an object-connotation theory, the authors argue in general for the cumulative-habit theory. They contend, rightly no doubt, that the nativist position cannot be adequately tested short of a longitudinal study of infants from birth. They believe that generalized past experience accounts for our major dispositions to perceive stroboscopic or other illusory movement, but allow that under conditions of ambiguity or equivocation specific meaning connotations will enter to determine the direction and nature of the movement. Here, too, our findings are concordant.

If we leave the field of experimental testing for a moment, we can find many familiar instances of the role of object connotation in resolving perceptual ambiguities. A streak of light in the night sky may be seen as a shooting star, as distant fireworks, or as a jet plane, depending largely on one's expectations. Bartlett tells of the Swazi chieftain who perceived all traffic policemen in London as friendly beings, because in Swazi culture the upraised arm is an amiable greeting. A child in a dentist's chair, more familiar with space-ships than with nitrous oxide, perceives the inhalator as a spaceship toy. Every projective test assumes that ambiguous (multivalent) stimuli will receive subjective structuring on the basis of need, set, expectancy, or habit.

Our experiment does not introduce factors of need or of set, but deals only with the relevance of past experience (meaning) as a determinant of perceived movement. It may, however, be pointed out that among sophisticated observers of the trapezoidal illusion under marginal conditions (e.g., at 10 feet binocularly) a volun-



tary effort to see or not to see the window as oscillating (or as rotating) can also be effective, especially if the observer picks out some detail of the window to watch during the rotation, thus inhibiting the impression as a whole. Meaning is not the only determinant entering into the resolution of perceptual ambiguity, but it is one of them.

Returning to James's statement that, "Perception is of definite and probable things," we may say that under optimal conditions of stimulation definite structure is conferred by physiological conditions or by deeply ingrained functional habits of spatial adjustment, or by both. But when marginal conditions prevail, an association with the most "probable" object is often called upon to provide the definiteness that is otherwise lacking.

What we have called "marginal" conditions should receive a further word of explanation. We use the term in our experiments to indicate that perceptual conflict is present. Under binocular conditions (especially at 10') there are many cues that "give away" the true rotation; at the same time there are operating also the assumptions that the window is rectangular and that longer objects are nearer. Under such a condition of conflict our finding is that urban children resolve the conflict with the aid of the supplementary assumption of "windowness." Not being able to draw on this supplementary assumption, the rural children as a rule resolve the conflict in favor of the binocular (or true) evidence. In this particular case, therefore, one might say that the primitive children see things "as they are" more often than do the children of civilization.

#### CONCLUSION

The perception of motion as represented in the rotating trapezoidal window is governed, under *optimal* conditions, by nativistic determinants or by the unconscious utilization of residual (but not immediately relevant) experience, or by both. (Our experiment does not enable us to decide this issue.) At the same time, object connotation (meaning) based on closely relevant cultural experience helps to determine the nature of the perceived movement under *marginal* conditions.

An adequate theory of perceived movement must therefore allow a place for the subject's specific assumptions of meaning even though it cannot be based solely on this foundation.

#### REFERENCES

1. ALLPORT, F. H. *Theories of perception and the concept of structure*. New York: John Wiley, 1955.
2. AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No. 7 (Whole No. 324).
3. BRENNER, MAY W. The effects of brain damage on the perception of apparent movement. *J. Pers.*, 1956, **25**, 202-212.
4. CANTRIL, H. *The "why" of man's experience*. New York: Macmillan, 1950.
5. DESILVA, H. R. An experimental investigation of the determinants of apparent visual movement. *Amer. J. Psychol.*, 1926, **37**, 469-501.
6. DOOB, L. An introduction to the psychology of acculturation. *J. soc. Psychol.*, 1957, in press.
7. HALL, K. R. L., & EARLE, A. E. A further study of the pendulum phenomenon. *Quart. J. exp. Psychol.*, 1954, **6**, 112-124.
8. HELMHOLTZ, H. VON. *Physiological optics*. Vol. III, Chaps. 26, 78. Transl. by J. P. C. Southall. Optical Society of America, 1925.
9. JAMES, W. *Principles of psychology*. Vol. II, Chap. 19. New York: Henry Holt, 1890.
10. JONES, E. E., & BRUNER, J. S. Expectancy in apparent visual movement. *Brit. J. Psychol.*, 1954, **46**, 157-65.
11. KILPATRICK, F. P. Two processes in perceptual learning. *J. exp. Psychol.*, 1954, **47**, 362-370.
12. KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
13. KORTE, A. Kinematoskopische Untersuchungen. *Z. f. Psychol.*, 1915, **72**, 193-206.
14. METZGER, W. *Gesetze des Sehens*. Frankfurt: Kramer, 1936.
15. MICHOTTE, A. *La perception de la causalité*. Paris: Vrin, 1946.
16. MÜLLER, G. E. *Komplextheorie und Gestalttheorie*. Göttingen: Vandenhoeck & Ruprecht, 1923.
17. NEFF, W. S. A critical investigation of the visual apprehension of movement. *Amer. J. Psychol.*, 1936, **48**, 1-42.
18. PASTORE, N. Some remarks on the Ames oscillatory effect. *Psychol. Rev.*, 1952, **59**, 319-323.
19. PIÉRON, H. Remarques sur la perception du mouvement apparent. *Année Psychol.*, 1934, **34**, 245-48.
20. RIVERS, W. H. R. *Cambridge anthropological expedition to Torres Straits*. Reports, Vol. II, Pt. 1: Introduction and Vision. Cambridge Univer. Press, 1901.
21. SCHILLER, P. VON. Stroboskopische Alternativversuch. *Psychol. Forsch.*, 1933, **17**, 179-214.
22. TOCH, H. H., & ITTELSON, W. H. The role of past experience in apparent movement: a revaluation. *Brit. J. Psychol.*, 1956, **47**, 3, 195-207.
23. WERTHEIMER, M. *Experimentelle Studien über das Sehen von Bewegung. Drei abhandlungen zur Gestalttheorie*, Erlangen: Verlag d. Phil. Akad., 1925, 1-105.
24. WUNDT, W. *Gründzüge der physiologischen Psychologie*, Vol. II, (6th ed.) Leipzig: Engelmann, 1910.

Received April 23, 1957. Prior publication.

# A MOTIVATIONAL FACTOR IN CONFORMITY<sup>1</sup>

GEORGE MOELLER AND MORTIMER H. APPLEZWEIG

Connecticut College

CONFORMITY in group situations may arise from many sources. This is a report of research into the effects on conformity behavior of individual motivation for social approval.

As we have indicated elsewhere (2, 3, 4), any of a number of behavior modes may serve as means toward satisfaction of a particular motive, and, conversely, a particular mode may serve any of a number of motives. This point of view implies that prediction of an individual's mode of behavior in a given situation must be based on knowledge of the situation and of the individual's motivation. Conforming behavior, as an example of one such mode, might then be predicted from a knowledge of individual motivation patterns, provided the performance situation permits gratification of the relevant motives in this way.

The present study was designed as a test of this general hypothesis. It was predicted that college students with high motivation for social approval would be more likely to agree with the judgments of their peers in the type of conformity situation developed by Asch (5, 6) than would students who were not so motivated. This hypothesis in no way precludes the possibility that conformity behavior may result from other patterns of motivation in other situations, but, rather, states a specific relationship between the strength of one motive and the occurrence of one mode of behavior in a particular situation.

## METHOD

### *The Measure of Motivation*

A new forced-choice motivation questionnaire, *The Behavior Interpretation Inventory* (9), was administered as part of the regular freshman orientation testing program at Connecticut College in the Fall of 1954. The inventory consists of 59 items, the stem of each repre-

sented a small sample of behavior, with four "reasons" given as alternatives. The four alternatives for each item are specific to the item, but are always some form of (a) escape from present unpleasantness, (b) concern for a secure future, (c) attainment of social approval, and (d) consistency with a self-picture. These four general motive categories were selected after a logical analysis of systems of classification of motives by Maslow (8), Murray (10), and Stein (15). A distinction was drawn between specific behaviors (modes) and generalized goal orientations (motives). Behaviors often grouped under the category of achievement motivation, for example, were considered to represent more appropriately behavior modes than underlying motives. A particular mode (e.g., acquisition, aggression, affiliation, etc.) can serve any one or more of an individual's motives, depending upon the situation in which it is called forth.

The utility of these particular motive categories was checked empirically in a study using an open-ended (sentence completion) form of *The Behavior Interpretation Inventory* (1). The four categories were found sufficiently inclusive to permit scoring of all response themes in the protocols obtained. Independent evidence for the usefulness of this motive classification system can be inferred from a recent study of men under military combat stress (11) where a list closely approximating our categories was obtained. The four choices likewise resemble the four "wishes" of Thomas (17).

Subjects (Ss) were asked to rank each set of four alternatives in the order of likelihood that they were the reasons for the behavior described in the item stem. Since the item stems represent a variety of situations and behavior modalities, the frequency of preference for a given motive was taken as an index of the strength of the motive relative to that of the other motives for a particular individual. Score on a given scale, then, was the total number of first choices assigned to alternatives appropriate to that scale.<sup>2</sup> A total of 25 behavior modes was used in the 59 items of the inventory, covering a wide range of situations and types of action. The two examples below represent modes of counteraction (#35) and physical sentence (#52). In brackets following each alternative is the number representing the motive category (see above).

35. He has been in poor health for some time, but, despite this, he is engaging in strenuous physical activity because:

(A) He fears his health might get even worse if he should take it easier [2].

(B) This relieves some of his tensions and lets him sleep better [1].

<sup>1</sup> This article constitutes Technical Report No. 5, Project NR 172-228, Contract Nonr 996(02) between the Group Psychology Branch, Office of Naval Research and Connecticut College. Permission is granted for reproduction in whole or in part for any purpose of the United States Government. A report of this research was given at the meetings of the Eastern Psychological Association, New York City, April, 1957.

<sup>2</sup> Alternative scoring for the sum of ranks has been used and is described elsewhere, as is the Inventory and its rationale in greater detail (2, 3, 4, 10, 16). The inventory was originally developed by the authors jointly with Drs. Dee G. Applezweig and Ross R. Thomas. Dr. Harvey Burdick participated in the revisions which resulted in the present form.



TABLE 1

MEAN BII  $z$ -SCORE PROFILES FOR SOCIAL-APPROVAL, SELF-APPROVAL AND SOCIAL- AND SELF-APPROVAL GROUPS

Group High	BII Scale:				
	<i>N</i>	1 Escape	2 Secu- rity	3 Soc. Appr.	4 Self Appr.
on					
Social Approval	15	-0.51	0.07	1.27	-0.51
Self Approval	13	-0.51	-0.73	-0.46	1.10
Social & Self Approval	13	-0.78	-0.56	0.37	0.67

(C) He has always believed in keeping one's body strong [4].

(D) He hopes that others will admire him for trying [3].

52. This man is lolling in the warmth of the water at a hot-springs resort. He often comes here because:

(A) He finds his mind is clear and he can think best when he is relaxed [4].

(B) He feels it is socially fashionable to vacation here [3].

(C) The sulphur of the water and the warmth relieve his rheumatism [1].

(D) He hopes that this will help him stay healthy longer [2].

### Subjects

BII scores of 263 *Ss* in the class were divided into four groups on the basis of scores on the scales of social and self approval (scales 3 and 4, respectively). Individuals were classified as high and low on each scale as their scores fell above or below the median on that scale. Scores on the two irrelevant scales (1 and 2) were used as a further restriction on sampling to produce groups with approximately equal means on these two scales. A test sample of 88 *Ss* was finally obtained, consisting of *Ss* high on social approval (scale 3) and low on self approval (scale 4), *Ss* high on self approval and low on social approval, and *Ss* high on both scales. Of the 88 *Ss* run in the experiment, 41 were used, the remainder being discarded for the following reasons: 3 for error in the conduct of the experiment, 6 for error on the part of a member of the majority, 10 who guessed the true purpose of the experiment while participating in it, and 28 who appeared on subsequent interview to have been overly suspicious of the procedure and/or purpose of the experiment.<sup>3</sup>

Table 1 presents mean BII  $z$ -score profiles for the three groups whose data compose this study. It should be noted that the group high on both scales is actually

<sup>3</sup> Interviews were tape-recorded and reviewed by the authors without knowledge of the motivation-group to which the *S* belonged. Conformity score was likewise ignored in determining whether the data for a particular *S* would be discarded or not. Actually, the conformity scores of the 38 *Ss* discarded (for knowledge or suspicion of the purpose of the experiment) varied considerably, although the mean number of yields for this group was smaller than for those *Ss* used.

not as high as either of the other groups on the relevant scale. This matter will be discussed further below.

### Procedure

The experimental procedure followed that employed by Asch in his studies of the effect of social pressure on the judgment of length of lines (5, 6), with the exception that the size of the majority used for the present study ranged from 4 to 6, in most cases being 5, while Asch used 6 to 8 member groups.<sup>4</sup>

Each *S* was asked to participate in a study of perception. On reporting, *S* found 4 to 6 of her classmates (who had ostensibly received identical requests) waiting for the experimenter (*E*). The group was brought to the experimental room and seated in a semicircle facing the blackboard. *E* then read the following instructions:

This is a task involving the discrimination of lengths of lines. Before you is a pair of cards. On the left is a card with one line; the card on the right has three lines differing in length; they are lettered A, B, and C, in order. One of the three lines at the right is equal to the standard line at the left—you will decide in each case which is the equal line. You will state your judgment in terms of the letter of the line. There will be 18 such comparisons in all.

As the number of comparisons is few and the group small, I will call upon each of you in turn to announce your judgments, which I shall record here on a prepared form. Please be as accurate as possible. Do not talk to each other during the experiment. Any questions? Suppose we start at the right (*Ss*' right) and proceed to the left around the table. Will you begin, please?

The experiment was then conducted in accordance with the instructions. The 4 to 6 classmates of *S* had been instructed beforehand to give specific erroneous responses on 12 of the 18 trials. To reduce the possibility of error by the majority, the first member was given a list of the correct responses which she consulted inconspicuously when necessary; all other members of the majority repeated the leader's response. The majority was instructed to respond as though their judgments were truly their own, but not to overact their parts. In all cases the members of the majority took their seats so that all but one member was to *S*'s right, the other to her left.

At the conclusion of the experiment, *E* asked the members of the group, in turn, to report to separate rooms for a brief interview. *S* was, of course, the only member of the group actually interviewed. This interview served to inform *S* of the true nature of the experiment and thus relieve some of the anxiety which may have been generated in the test situation. It also provided information regarding the set *S* had taken, particularly whether or not she had accepted the instructions and the experiment at face value. The authors alternated as interviewer as their schedules permitted.

<sup>4</sup> Mrs. Joan Lake Kaiser served as experimenter in the conformity measurement procedures. We wish to express our thanks to her and to the many students who served as experimental assistants and subjects.

## RESULTS

*Quantitative*

Figure 1 shows the percentage of Ss in each group which yielded to the majority on the twelve trials in which the majority responded incorrectly. It can be seen that the social-approval group consistently yielded more frequently than did either of the other two groups. The performance curves of the self-approval group and the group high on both scales could not be distinguished from each other. Further, from the third critical trial forward, the performance of the three groups seemed to reach a stable level, such differences as appeared neither increasing nor decreasing consistently.

Table 2 shows the number of Ss who yielded to, or compromised with, the majority a given number of times for each of the three groups. It can be seen that the frequency distribution for the social-approval group is somewhat skewed toward high frequency of yielding; while the distributions for the other two groups are skewed in the opposite direction. The Kruskal-Wallis nonparametric measure (H) for comparing sum of ranks (14) yielded a

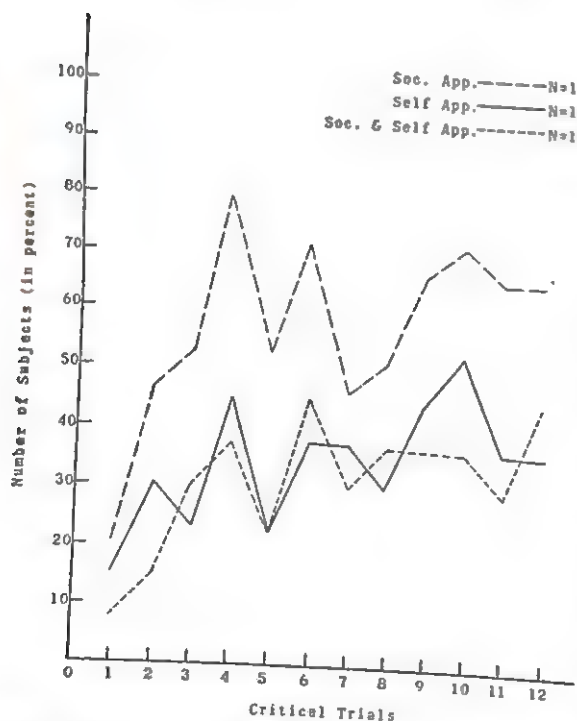


FIG. 1. PERCENTAGE OF CONFORMING AND COMPROMISE RESPONSES FOR EACH PROFILE GROUP ON THE TWELVE CRITICAL TRIALS

TABLE 2  
FREQUENCY OF COMPROMISE OR YIELDING IN EACH OF THE EXPERIMENTAL GROUPS ON THE TWELVE CRITICAL TRIALS

Frequency	Experimental Groups		
	Soc. Appr. (Scale 3)	Soc. & Self Appr. (Scales 3 & 4)	Self Appr. (Scale 4)
	N = 15	N = 13	N = 13
12	1	—	1
11	1	2	—
10	1	1	—
9	2	—	—
8	3	—	2
7	1	—	1
6	2	1	1
5	1	—	2
4	—	2	—
3	1	1	—
2	1	—	2
1	1	1	—
0	—	5	4

$p$  close to .05 for these data, indicating that the groups do in fact differ in the predicted direction. Comparison of the several pairs of treatments by  $U$  tests further showed that the social-approval group differed from the other two,  $p$  being less than .05 in both cases (single-tailed test). The other two groups do not differ from each other.

*Qualitative*

The postexperimental interview yielded qualitative data that generally support the findings reported for the experiment proper. The standard questions that elicited different responses from the several groups, or that are of interest in themselves, are shown in Table 3, along with the frequency of the various responses. Data for the two Ss who always agreed with the majority are not included since most of the questions could not be asked of them.

Only Questions 1 and 7c evoked answers requiring any kind of clinical evaluation in scoring. These answers were typed verbatim, then analyzed without reference to the Ss' group membership. The criteria for classification are presented below along with appraisal of the answers. In no case did the analysis go beyond the explicit content of Ss' statements.

Responses to the first question were classified by the authors as intro-punitive or nonintro-punitive, as this concept is defined by Rosenzweig (12). Intro-punitive responses were those in which S expressed doubt about the accuracy



TABLE 3

## RESPONSES IN THE POST-EXPERIMENTAL INTERVIEW

## Question 1:

Would you describe in your own words your experiences during this experiment?

Response Classified as:	Group		
	Soc. App.	Soc. & Self	Self App.
Intropunitive	8	3	5
Nonintropunitive	6	10	7

## Question 2a:

When you gave an estimate that disagreed with the others, did you feel that if the lines were measured with a ruler you would turn out to be right or wrong?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Right	10	12	12
Wrong or ?	4	1	0

## Question 2b:

Then did you feel that if the lines were measured, the others would turn out to be wrong?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Wrong ? or Right	7	12	10
	7	1	2

## Question 3:

Would you say that you were concerned about the disagreements?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
No	5	8	7
Yes or ?	9	5	5

## Question 4a:

Still thinking back to the time that you were giving your judgments, would you say that the others made you doubtful about your accuracy?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Yes	11	7	5
No	3	6	7

TABLE 3—Concluded

## Question 5a:

Would you say that you were tempted at times to answer as the others did?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Yes	12	7	5
No	2	6	7

## Question 6a:

Did you ever answer as the others did against your own first choice?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Yes	7	4	2
No	7	9	10

## Question 7b:

Did you wonder what everybody might be thinking of you when you disagreed?

Response	Group		
	Soc. App.	Soc. & Self	Self App.
Yes	10	9	9
No	4	4	3

## Question 7c:

What did you think the others thought about you?

Response Classified as:	Group		
	Soc. App.	Soc. & Self	Self App.
Something wrong with me	7	6	1
Other	3	3	8

of her judgment, e.g., "I think when she took them away I kept wondering, did I answer correctly . . . they were so close sometimes." A response was classified as nonintropunitive if S stated that the group was in error, or that some physical aspect of the situation led to the disagreements, or she simply reported that there were disagreements. An example of the latter type of response is, "Everybody was saying the line was A, and I said it was B. No-one else ever varied from their observations, but I did." The content of these examples is

typical of those assigned to the two categories, but they are unusually brief.

The responses to Question 1 of the social and self-approval group were more often non-intropunitive than were those of the other two groups. In response to Questions 2a to 6a, the social-approval group differed in the expected way from the other two groups. Members of that group were: less certain that they were right (Q 2a), less willing to state that the majority was wrong (Q 2b), more concerned about disagreements with the majority (Q 3), less certain about the accuracy of their judgments (Q 4a), more often tempted to answer as the others did (Q 5a), and more willing to admit that they had answered against their own first choice (Q 6a). None of these differences is completely convincing when taken alone, but in the aggregate they are consistent with the hypothesis that members of the social-approval group were more concerned over their performance relative to that of the majority than were members of the other two groups.

The differences that appeared in the interviews do not simply reflect differences in performance in the experiment proper. If we compare the performance scores of persons who said "Yes" and "No" (or "Right" and "Wrong") in reply to a particular question we find no differences. This lack of relationship between performance and *S*'s reaction to that performance arises, of course, because *S* may yield uncritically, yield but feel concern, not yield and feel concern, or remain independent in both performance and attitude.

The responses to Questions 7b and 7c clearly reflect this distinction between behavior and the concomitant affect. Data for Q 7b are presented to show that awareness of others was about equally characteristic of all groups. The responses of those who replied "Yes" to Question 7b are analyzed for Question 7c.

Responses to Question 7c were classified as "Something wrong with me" or "Other." Of those who stated that there was "Something wrong with me," nine persons said something like, "I thought maybe they were thinking I was just crazy"; the other five said, "I thought they might be wondering if I was half blind or something," or its equivalent. Members of the "Other" group either equivocated, "I was just wondering what they were thinking," or,

simply accepted the fact, "I thought they probably thought I was wrong."

The responses to Question 7c provide further evidence regarding the difference between the social- and self-approval group and the self-approval group. The replies of the group high on both scales clearly resemble those of the social-approval *S*s and not at all those of the self-approval *S*s.

## DISCUSSION

Figure 1 and Table 2 show that the social-approval group did, as predicted, yield to an erring majority more frequently than did groups of *S*s with other motivational profiles. The qualitative data, Questions 2a to 6a, are also in accord with the prediction. It should be emphasized that our prediction was not based on categorization of persons as conformers and nonconformers. As we have previously indicated (2, 3, 4), we assume that conformity *per se* is not a trans-situational trait, but that any individual conforms or fails to conform in a given situation depending upon his motivational profile and the relevance of conformity behavior in that situation to his motives. Thus, in a situation in which no motive gratification is to be expected from conforming, such behavior would not be predicted even where there is reason to believe social approval is an important motive in the individual's life. In an unpublished study from the authors' laboratories,<sup>5</sup> for example, the following interesting results were found. In an Asch judgmental situation similar to the one used in this study, greater conformity was obtained when the majority consisted of a high prestige group (college seniors) than when it was composed of peers (college freshmen). Even more striking was the finding that the presence of a low prestige majority (high school sophomores) decreased the number of yielding responses and increased the number of errors in the direction *opposite* to that taken by the majority. These findings are considered to be entirely consistent with the present interpretation of conformity behavior as instrumental activity.

The distribution of conformity scores for the social- and self-approval group shows a tendency toward bimodality, but the number

<sup>5</sup> This study was conducted by Misses Jane Molloy and Barbara Munger as an undergraduate course research assignment.



of cases is too small for this departure from unimodality to be significant. Such bimodality might suggest that individuals in this group choose one resolution or another of the motive conflict, i.e., they conformed or did not. The qualitative data are quite interesting in this regard.

The answers to Question 1 do little more than suggest that the social- and self-approval group is in some way different from the others. At that stage of the interview, most of the Ss gave rather evasive answers which they contradicted later when asked about specific aspects of the experiment. For instance, five members of the self-approval group gave seemingly intro-punitive responses to Question 1, but all members of that group stated that they believed they had given accurate judgments in response to Question 2a.

The data presented for question 7c in Table 3 in conjunction with the quantitative measures seem to establish the status of the social- and self-approval group. It appears that this group may have been in conflict, as we expected, a conflict which they may have resolved by differentiating between their performance and their evaluation of the majority's reaction to it. We noted the apparent lack of relationship between performance and attitude previously.

The studies presenting clear evidence of conflict in performance involve specific motor acts (7). Evidence of other conflicts, however, is typically clinical in nature (13). The implication of the literature and of our study seems to be that in the study of conflict in complex situations simple performance measures must be supplemented by other means. It appears that the social-approval Ss yielded uncritically, or yielded but felt concern; the self-approval Ss remained independent in performance and attitude; and the social- and self-approval Ss did not yield, but felt concern over the majority's opinion of their independence.

The factors discussed above appear to account for the similarities and differences between the motivational groups. However, the final samples were not perfectly matched in all respects in which matching was attempted (see Table 1). The social- and self-approval group actually occupies positions intermediate between those of the social-approval group and the self-approval group on both scales 3 and 4.

This fact may have contributed to the absence of more clear-cut differences between the social- and self-approval group and the others. Second, the scale 2 (security) scores of the social-approval group are higher than those of the other groups. In order to test the effect of scale 2 scores, the 41 cases were divided into those above and below the median scale 2 score for the total sample. The conformity scores of the two groups so constructed were then compared by means of the *U* test, with the finding that scale 2 scores were not related to conformity behavior in our test situation.<sup>6</sup>

#### SUMMARY AND CONCLUSIONS

Two groups of women's college students were placed in a judgmental conformity situation of the sort studied by Asch to test the hypothesis that persons with high social-approval and low self-approval motivation would yield to a unanimous, but erring, majority in judgment of length of lines more frequently than would persons with the opposite motivational profile. A third group with high motivation in both areas was included to provide information about the effects of motivational conflicts in a conformity situation. The measures of social and self approval were obtained from a four-scale measure of motivation, *The Behavior Interpretation Inventory*.

The results were in accord with the major hypothesis. The groups differed as predicted in frequency of conformity to the erroneous judgments of the majority. Comparison of the groups by pairs indicated that the high social-approval—low self-approval Ss yielded to the majority significantly more often than did either of the other groups. The performances of the other two groups were similar.

Qualitative data obtained in postexperimental interviews indicated that Ss in the conflict group differed from those in the self-approval group in their evaluation of the majority's attitude toward their performances. It is suggested that the high social- and self-approval Ss experienced conflict, which they resolved by dissociating their performance from their emotional reactions to it.

<sup>6</sup> We do not mean to imply that concern for security (nonsocial) is necessarily unrelated to conformity behavior in all situations, but that this particular situation does not appear to have been influenced by this motive as measured.

## REFERENCES

1. APPLEZWEIG, DEE G. *Scoring the open-ended form of The Behavior Interpretation Inventory*. Technical Report, ONR Contract 996 (02), NR 172-228, Connecticut College, in press.
2. APPLEZWEIG, DEE G., & APPLEZWEIG, M. H. *Stress and behavior. I. The Behavior Interpretation Inventory*. Annual Technical Report, ONR Contract Nonr 996(02), NR 172-228, Connecticut College, Aug., 1954.
3. APPLEZWEIG, M. H., & MOELLER, G. *Motivation and psychological stress: Second annual report*. ONR Contract 996 (02), NR 172-228, Connecticut College, Feb., 1956.
4. APPLEZWEIG, M. H., MOELLER, G., & BURDICK, H. Multi-motive prediction of academic success. *Psychol. Reports*, 1956, 2, 489-496.
5. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
6. ASCH, S. E. Studies of independence and conformity: 1. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, 70, No. 9 (Whole No. 416).
7. HOVLAND, C. I., & SEARS, R. R. Experiments on motor conflict: I. Types of conflict and their modes of resolution. *J. exper. Psychol.*, 1938, 23, 477-492.
8. MASLOW, A. H. "Higher" and "lower" needs. *J. Psychol.*, 1948, 25, 433-436.
9. MOELLER, G., & APPLEZWEIG, M. H. *The Behavior Interpretation Inventory: Manual*. Technical Report No. 2, ONR Contract Nonr 996(02), NR 172-228, Connecticut College, May, 1956.
10. MURRAY, H. A. et al. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
11. OPERATIONS RESEARCH OFFICE, et al. *A study of combat stress Korea 1952 (Preliminary Report)*. Tech. Memo. ORO-T-41 (FEC), Project DAS-6 FAST, Operations Res. Office, The Johns Hopkins Univer., Chevy Chase, Md., Dec., 1952.
12. ROSENZWEIG, S. The picture-association method and its application in a study of reactions to frustration. *J. Pers.*, 1945, 14, 3-23.
13. SEARS, R. R., HOVLAND, C. I., & MILLER, N. E. Minor studies of aggression: I. Measurement of aggressive behavior. *J. Psychol.*, 1940, 9, 275-295.
14. SIEGAL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
15. STEIN, M. I. *The Thematic Apperception Test: an introductory manual for its clinical use with adult males*. Cambridge: Addison-Wesley Press, 1948.
16. THOMAS, R. R. *The use of structured material in projective testing*. Technical Report No. 1, ONR Contract Nonr 996(02), NR 172-228, Connecticut College, Sept., 1955.
17. THOMAS, W. I. *The unadjusted girl*. Boston: Little, Brown, 1923.

Received February 15, 1957. Prior publication.



# GROUP LEARNING AND COMMUNICATION AS A FUNCTION OF TASK AND STRUCTURE "DEMANDS"<sup>1</sup>

JOHN T. LANZETTA

*Fels Group Dynamics Center, University of Delaware*

AND THORNTON B. ROBY

*Institute for Applied Experimental Psychology, Tufts University*

SINCE Bavelas' early papers on group structure (1, 2) a number of experiments have examined the relationships between certain of his suggested communication structure properties and group performance (e.g., 3, 4, 6, 12). In general, the experiments have employed a self-paced task with a single decision required and have utilized nets differing in the number and pattern of communication links between subjects. As Luce *et al.* have noted, "... the prototype groups are those formed by society to produce accurate information outputs in a short period of time from certain input data—information processing groups" (8, p. 82).

The present experiment is an extension of several earlier studies (7, 9, 10, 11) which, although concerned with information processing groups, focused on a different characteristic of communication structure and employed a forced-paced task. The different emphases reflected an interest in problems of division of labor in a slightly different class of information handling groups, i.e., those formed to produce multiple outputs from certain classes of changing input data, where both output and input requirements entail some division of responsibility. One critical variable manipulated when allocations of responsibilities are varied is the degree of information exchange required between group members. The term "communication structure" was used to refer to the pattern and volume of required intermember communications which resulted from the nature of the task imposed on the group and the way responsibilities were allocated within the group (9). By imposing a task with fixed, and known, display-control relationships and manipulating the allocation of observation and

control action responsibilities within the group, structures were produced differing in terms of the volume and distribution of required intragroup communications.

As expected, structure difficulty decreased as the "autonomy" of group members increased, i.e., each member was the primary source of information required for his decisions, and information not directly available was located at a single external source (10, 11). That is, group efficiency decreased with an increase in the required amount of information exchange and with an increase in the amount of information exchange involving multiple sources. When the information processing demands of the task were increased, the relative differences between structures were maintained with errors for all structures increasing monotonically with increases in the amount of information presented the group per unit time. However, even under the "easiest" experimental conditions—with a low rate of information input and low information exchange structure—the groups were able to process (without error) considerable less information than might be expected from a knowledge of the information handling capacity of individuals (7).

Although the results supported the initial general hypothesis that structure-imposed demands for intragroup communication would be related in a systematic way to group performance over a range of task imposed demands, they at the same time raised additional issues. Three such interrelated issues motivated the present experiment.

1. There was some evidence that learning occurred under all of the task and structure conditions explored, but it was not possible to determine whether the learning curves were marked by sharp discontinuities,<sup>2</sup> and whether

<sup>2</sup> Group "learning" curves probably reflect the operation of two primary factors: acquisition and improvement of relevant part skills of group members and improvement in organization or "coordination." Pre-

<sup>1</sup> This research was carried out under the Air Force Personnel and Training Research Center, Lackland Airforce Base, San Antonio, Texas, in support of Project 7713. Permission is granted for reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government.

there were qualitative and/or quantitative differences in the curves as a function of task input and structure "demands."

2. In addition, if improvement in performance occurred, it should be accompanied by changes in the communication behavior of group members. If so, would the changes in communication be a function of the requirements for information exchange imposed by structure and input variations, and could the performance of groups be related to the adequacy with which they adapted their communication behavior to these imposed "demands"?

3. Another aspect of the problem of "adaptation" was the question of the effects of group operating procedures for exchanging information on performance. As we noted above, the information processing capability of the groups was lower than might be anticipated from data on individual performance. Such a discrepancy may have resulted from the method of varying information adopted from the experiments and the task used. However, there was some indirect evidence that the primary limiting factor was the inability of groups to set up an efficient system for information exchange (7, 11).

Groups appeared to be using a "soliciting" rather than a "volunteering" system: The individual who required information called the person who was a source and requested the information rather than the source person calling the user when a change of information had occurred. Since the latter procedure entails just half the number of messages as the former, it might be expected that information transmission would be more effective using a "volunteering" system. However, volunteered information may reach a group member "out of phase" with other essential information (and would have to be stored) or at a time

sumably, the form of the learning curve in any particular case will depend on the relative contribution of part skills and coordination to the group product measure. In general, however, the acquisition of part skills should produce gradual improvement in efficiency, while adoption of a new procedure or organization could produce a more marked change in efficiency. Thus, a learning curve based on group product measures may be of typical exponential form but with sharp changes in slope at various points, the discontinuities being associated with the adoption of a new procedure or reflecting a reorganization of the group. Data from the early experiments had not revealed such discontinuities, but very few trials had been used.

when other activities conflict with attending to the message (and may be lost). Thus, despite the fact that fewer messages are required, volunteering could prove to be a less effective procedure than soliciting. Data on the relative efficiency of volunteering and soliciting under a range of input and structure conditions would help clarify the issue.

The present study then, is an extension of previous experiments and is designed to throw some light on the issues raised above. Since a scarcity of data bearing on group learning or communication is available in situations of the sort considered here, no attempt is made to advance or test specific hypotheses. Rather, the focus is on obtaining objective measures that are extensive and reliable, in order to shed some light on the relationships between task and structure "demands" and group performance.

## METHOD

### Task

The task used in this study was essentially that described in previous reports (7, 10). There were three major components, the displays, the controls, and the communication system.

The displays were 6 AC voltmeters on each of which an aircraft "instrument" dial was simulated. The dial readings were programmed by simple resistance switching accomplished at a central console. Each instrument could be set at one of three positions. The controls were 6 two-position switches which were also labeled to simulate aircraft controls. Associated with each switch was an "operating procedure" card which indicated the correct switch settings for the various values of instrument readings. Correct settings for each switch were determined by the readings on a particular pair of instruments.

Three member groups were used. Each subject (*S*) was assigned to a three-sided booth in which was placed two simulated aircraft instruments, two control switches, and a "communication signalling" box (Fig. 1). Direct communication was prevented. The *Ss* could communicate (over a standard 6-volt interphone circuit with conventional headsets and microphones) by depressing switches which provided communication connections between pairs of booths. The signalling device in conjunction with an operations recorder provided an automatic and continuous record of "who called whom" and the length of time spent in each communication. The settings of each of the six control switches were also recorded on a 20-pen Esterline-Angus operations recorder.

The task of the group was to adjust and readjust each of the six controls to their proper settings in response to periodically changing readings on the instruments. The instruments in each booth contained only part, or none, of the information required for adjusting the controls in that booth, the additional information



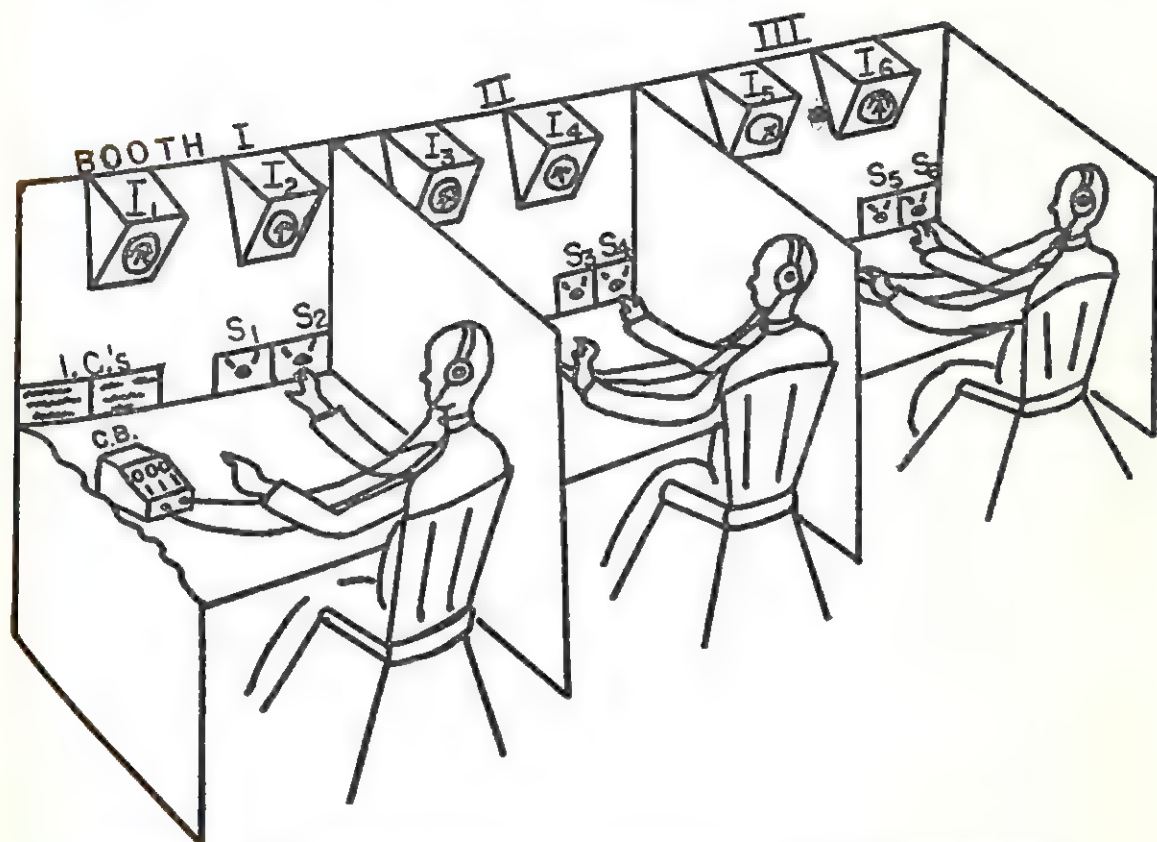


FIG 1. SKETCH OF APPARATUS AND EXPERIMENTAL ARRANGEMENT: I 1-6 = INSTRUMENTS; S 1-6 = SWITCHES; I C's = INSTRUCTION CARDS; C.B.'s = CALL BOXES

being located in one or both of the other booths. Each *S*, then, functioned as both a control agent and an information source person with responsibility for adjusting two controls and detecting and relaying instrument reading changes to other *S*s.

### Experimental Conditions

**Structure.** Two structure conditions were employed. In the Low Autonomy Structure (LA) each *S* had two controls for which none of the four necessary instrument readings was directly available. In High Autonomy Structure (HA) each *S* had two controls for which all but one of the requisite instrument readings were available in his booth. Thus, in the LA structure, each *S* was dependent upon the other group members for all his required information, while in the HA structure each *S* was dependent upon external sources for only one reading. The interdependencies between group members represented by a matrix schema is shown in Table 1.

**Task load.** Task load was defined as the rate of change of the instrument readings. The input to the group consisted of the readings on each of six instruments. At a fixed interval the readings on one of the six instruments changed requiring the adjustment of either one or two controls. A trial consisted of 18 such changes, three for each instrument. The sequence of instrument reading changes was quasi-random, subject to the restriction that there be three changes for each

instrument. The number of control adjustments that each of the three *S*s had to make in response to the 18 changes was equalized.

Three fixed intervals between instrument reading changes were employed: 6 seconds, 10 seconds, and 15 seconds. Since the number of changes was constant for a trial, this necessitated having trials of unequal length for the three different "loads."

**Operating procedure.** As noted above, one comparison of interest in this study was between a "volunteering" and "soliciting" procedure for information exchange. Complete control over the procedure used by the groups was unfeasible, but preliminary experience indicated that a large measure of control could be obtained by differential instructions. The initial instructions to groups assigned to the volunteering procedure stressed the role of the *S* as an information source. The *S*s were told who needed the information they had, and were asked to report any changes in their readings to the group members who required the information. In contrast, the instructions to "soliciting" groups emphasized the responsibilities of each *S* as a control agent. The *S*s were told which other group members had the information required for correctly adjusting their controls and were asked to call these *S*s to obtain the information. In both cases, several practice trials were performed under experimenters' supervision in which the appropriate procedure was again stressed.

TABLE 1  
INFORMATION TRANSMISSION REQUIREMENTS IN THE  
TWO EXPERIMENTAL STRUCTURES\*

	Group Member	Source of Information		
		A	B	C
High Autonomy Structure (HA)				
User of information	A	3	1	0
	B	0	3	1
	C	1	0	3
Low Autonomy Structure (LA)				
User of information	A	0	1	3
	B	3	0	1
	C	1	3	0

\* The entries indicate the items of information (instrument readings) held by the group member indicated in the column and required by the individual indicated in the row. Entries in the diagonal indicate items directly available.

### Subjects and Design

The Ss were basic trainees at Lackland Air Force Base. Six Ss reported for testing each day; they were divided fortuitously into two teams of three Ss each for morning and afternoon testing. They were brought into the experimental room, seated in their booths, and oriented as to the general nature of the task and the use of the interphone equipment. Following this, they were given a brief practice session to insure that they understood how to interpret the instrument readings and the operating instruction cards. They then performed the task under one of the combinations of the experimental conditions.

There were three experimental variables: Structure, Operating Procedure, and Load, with two conditions of structure and procedure and three levels of load. A  $2 \times 2 \times 3$  factorial design was used with four groups randomly assigned to each cell. Each group performed nine trials, each trial consisting of 18 instrument reading changes. After each set of three trials there was a short two- to three-minute rest period during which Ss remained in the booths.

### RESULTS<sup>3</sup>

The scores used in the analysis are the number of errors in control settings made by the group within each of the nine trials. A change in an instrument reading in any of the three booths required either one or two control adjustments; if the correct control settings were not made before the next change in an instrument reading an error was recorded for each incorrect setting.

Table 2 presents the error scores for the

<sup>3</sup> Acknowledgment is made for the able assistance of Mr. Edward Amberg both in experimental administration and in the statistical analysis reported.

experimental conditions and trials. The analysis of variance of these data is shown in Table 3.

Congruent with the results of earlier studies, structure and load are significant. Errors increased as the rate of change of information increased and were greater for the LA structure. The difference between the "volunteering" and "soliciting" procedures is not significant, although in the expected direction (lower errors for "volunteering").

The load  $\times$  structure interaction is significant when tested against error (b) but not when tested against the significant second order interaction—load  $\times$  structure  $\times$  procedure. In general, the HA structure was superior to the LA structure, the difference being most pronounced for the 10-second rate of change. However, the relative superiority of HA over LA was also a function of the procedure used; a volunteering procedure resulted in a greater difference between HA and LA for the 6-second rate of change, while the converse was true for the 15-second rate of change.

Although the six conditions of structure and load cannot unequivocally be ordered in terms of difficulty, there is little doubt that the 6"-LA, and 15"-HA conditions, respectively, define the most and least difficult ends of the continuum. Thus, volunteering appears to be more effective than soliciting for intermediate difficulty levels.

Other sources of significant variance are those due to trials and the interactions of trials  $\times$  load and trials  $\times$  structure. Examination of the over-all means in Table 3 indicates that there was a monotonic decrease in error scores as a function of practice. Inspection of Fig. 2 and Fig. 3 indicates, however, that the regression of errors on trials was a function of both load and structure. The rate of improvement increased as the rate of change of information decreased, and was greater for the HA structure and the LA structure. There was some evidence of a performance asymptote in the 15-second trend and in the trends for the LA and HA means.

*Analysis of communication records.* From the 20-pen Operations Recorder record of the "calls" between group members, three inter-related measures were obtained for each trial: The number of messages initiated by all three of the group members, the average length of



TABLE 2

MEAN ERRORS FOR EXPERIMENTAL CONDITIONS AND TRIALS: ENTRIES ARE THE MEAN OF ERRORS FOR FOUR GROUPS

FOUR GROUPS												
Procedure	Structure	Load	Trials									
			1	2	3	4	5	6	7	8	9	
Volunteering	LA	6"	49	49	44	47	49	46	46	52	47	429
		10"	55	38	42	41	38	39	38	34	29	354
		15"	36	33	26	20	18	20	19	18	23	213
			140	120	112	108	105	105	103	104	99	996
	HA	6"	46	42	43	37	39	34	30	23	34	328
		10"	30	21	21	11	15	13	11	10	7	139
15"		36	25	21	17	21	20	16	16	15	187	
		112	88	85	65	75	67	57	49	56	654	
Soliciting	LA	6"	49	48	48	47	41	44	48	49	41	415
		10"	53	49	45	46	39	37	36	42	40	387
		15"	41	38	33	34	30	31	30	27	26	290
			143	135	136	127	110	112	114	118	107	1092
	HA	6"	47	46	40	39	39	36	35	35	35	352
		10"	36	30	23	23	17	17	16	17	15	194
15"		23	17	15	15	10	12	11	8	9	120	
		106	93	78	77	66	65	62	60	59	666	

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF ERROR SCORES

Source of Variance	df	Sums of Squares	Mean Squares	F Ratio
Total	431	86,705.		
A. Between groups	47	64,478.	1,372.	23.66**
Load	2	28,949.	14,474.	81.31**
Structure	1	21,832.	21,832.	122.65**
Procedure	1	434.	434.	2.44
L × S	2	3,917.	1,958.	11.00**
L × P	2	454.	227.	1.27
S × P	1	253.	253.	1.42
L × S × P	2	2,245.	1,122.	6.30**
error (b)	36	6,394.	178.	
B. Within groups	384	22,227.	58.	
Trials	8	9,168.	1,146.	35.81**
T × L	16	921.	58.	1.81*
T × S	8	597.	75.	2.34*
T × P	8	370.	46.	1.44
T × L × S	16	841.	53.	1.65
T × L × P	16	589.	37.	1.16
T × S × P	8	92.	12.	.38
T × L × S × P	16	480.	30.	.94
error (w)	288	9,169.	32.	

\* Significant at the .05 level.

\*\* Significant at the .01 level.

messages was largely restricted to reports of instrument readings.

As with the error scores we were interested in differences in these measures of communication behavior as a function of experimental conditions, as well as changes as a function of practice. In addition, however, it was of interest to know whether there were differences between "good" and "poor" performance groups, within an experimental treatment. For the latter objective, the four groups within a treatment were dichotomized on the basis of mean error scores over the nine trials and "performance" was used as the fifth classification variable in the analyses of variance. The summary tables of the analyses of variance for number of messages, average length of messages, and total talking time, are presented in Tables 4, 5, and 6.

*Number of messages.* For the "number of messages" measure, load, structure, and trials contributed significant variance. The number of messages increased as the rate of change of information decreased, was greater for the LA structure, and increased monotonically with practice. There were fewer messages with the volunteering procedure than with the soliciting procedure, although the difference was not significant.

messages, and the total time spent by all three group members in communication. It was expected that any change in communication behavior would be in terms of the frequency and duration of messages rather than in message content, since the content of

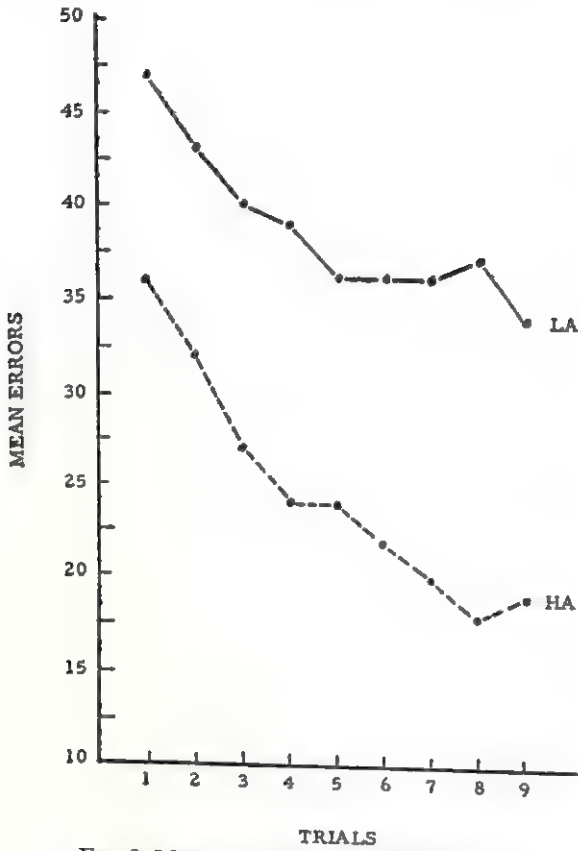


FIG. 2. MEAN GROUP ERROR SCORES OVER TRIALS FOR THE HA AND LA GROUPS

With respect to the result for the load variable, however, it was apparent that although the number of calls per trial increased as the rate of change of information decreased, the increase was proportional to the length of the trial. When corrected for length of trial, the number of messages per unit time for the three rates of change was approximately equal. The rate of message transmission appeared to be independent of the rate of change of input information; the differences between loads is attributable solely to the length of the trial.

*Average length of messages.* Table 5 shows that significant sources of variance for this measure are structure, performance, trials, and the trials  $\times$  load  $\times$  structure interaction. The average length of messages was longer for the low autonomy structure, shorter for the "better" performing groups, and decreased with practice. The changes with practice, however, were a function of both load and structure. For the LA structure the decrease was largest for the 10-second load and about equal for the 6- and 15-second loads, while for the

HA structure the largest decrease was for the 15-second load with the 6-second and 10-second loads following in that order. The amount of decrease, however, was correlated with the average length of messages on trial 1 and very little difference in mean message time was evident on trial 9 or in mean message time over the nine trials. The interaction may thus be reflecting a "catching up" phenomena which was accomplished by about the fourth trial.

The surprising feature of the results was the lack of significant variance in mean message time attributable to load. Since the amount of information to be transmitted per trial was constant for the three-load conditions, one might expect that decreasing the time for communication would have decreased the average length of the communications, but this did not occur. Apparently, as the volume of information to be transmitted increased (due to structure variations) the average length of messages increased, but there was no change in average length with changes in the amount of information to be transmitted per unit time (due to load).

*Total talking time.* As Table 6 indicates, load,

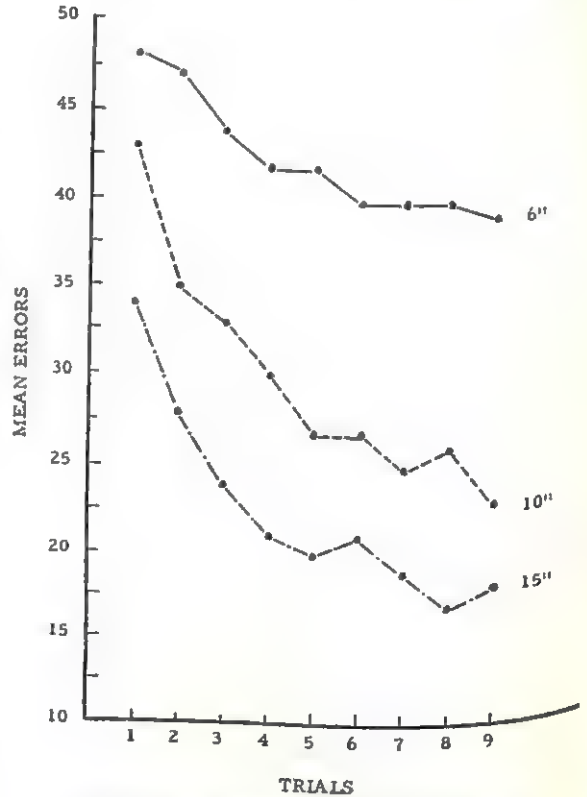


FIG. 3. MEAN GROUP ERROR SCORES OVER TRIALS FOR THE THREE LOAD CONDITIONS



structure, procedure, trials and the interactions—load  $\times$  structure, and trials  $\times$  structure—contributed significant variance to total talking time. Total talking time increased with a decrease in the rate of change of information, was greater for the LA structure than the HA structure, and was less for the volunteering than the soliciting procedure. Also, the difference between LA and HA structures, on this measure, increased as the rate of change of information decreased, e.g., the difference between LA and HA was greater for the 15-second load and the 6-second load.

As with the "number of messages" measure, when we corrected total talking time for the length of trials there were no differences between loads. It seems that whereas the volume of information to be relayed (structure) and the procedure used for routing information

TABLE 4  
SUMMARY OF ANALYSIS OF VARIANCE OF  
"NUMBER OF COMMUNICATIONS"

Source of Variance	df	Sums of Squares	Mean Squares	F Ratio
Total	431	170,917.		
A. Between groups	47	141,898.	3,019.	23.33**
1. Load	2	68,790.	34,395.	8.47**
2. Structure	1	12,392.	12,492.	2.66**
3. Procedure	1	3,918.	3,918.	—
4. Performance	1	1,306.	1,306.	—
1 $\times$ 2	2	2,314.	1,157.	—
1 $\times$ 3	2	1,888.	944.	—
1 $\times$ 4	2	2,276.	1,138.	—
2 $\times$ 3	1	5,132.	5,132.	3.48
2 $\times$ 4	1	65.	65.	—
3 $\times$ 4	1	6.	6.	—
1 $\times$ 2 $\times$ 3	2	3,340.	1,670.	1.13
1 $\times$ 2 $\times$ 4	2	906.	453.	—
1 $\times$ 3 $\times$ 4	2	1,499.	750.	—
2 $\times$ 3 $\times$ 4	1	643.	643.	—
Error (b) including quad. interaction	26	38,323.	1,474.	—
B. Within groups	384	29,019.	76.	22.94**
Trials	8	9,542.	1,193.	1.31
T $\times$ 1	16	1,090.	68.	1.04
T $\times$ 2	8	434.	54.	—
T $\times$ 3	8	364.	46.	—
T $\times$ 4	8	159.	20.	—
T $\times$ 1 $\times$ 2	16	829.	52.	1.00
T $\times$ 1 $\times$ 3	16	1,092.	68.	1.31
T $\times$ 1 $\times$ 4	16	359.	22.	—
T $\times$ 2 $\times$ 3	8	805.	101.	1.94
T $\times$ 2 $\times$ 4	8	134.	17.	—
T $\times$ 3 $\times$ 4	8	394.	49.	—
Error (w) including higher interactions	264	13,817.	52.	—

— F less than 1.  
 \* Significant at .05 level.  
 \*\* Significant at .01 level.  
 \*\*\* Significant at .001 level.

TABLE 5  
SUMMARY OF ANALYSIS OF VARIANCE OF "AVERAGE  
LENGTH OF COMMUNICATION"

Source of Variation	df	Sums of Squares	Mean Squares	F Ratio
Total	431	1,709,733.	—	—
A. Between groups	47	1,364,473.	29,031.	—
1. Load	2	950.	475.	—
2. Structure	1	749,750.	749,750.	56.99***
3. Procedure	1	17,088.	17,088.	1.30
4. Performance	1	92,138.	92,138.	7.00*
1 $\times$ 2	2	14,995.	7,498.	—
1 $\times$ 3	2	31,525.	15,762.	1.20
1 $\times$ 4	2	14,541.	7,270.	—
2 $\times$ 3	1	1,576.	1,576.	—
2 $\times$ 4	1	4,300.	4,300.	—
3 $\times$ 4	1	9,343.	9,343.	—
1 $\times$ 2 $\times$ 3	2	11,389.	5,694.	—
1 $\times$ 2 $\times$ 4	2	3,837.	1,918.	—
2 $\times$ 3 $\times$ 4	1	15,588.	15,588.	1.18
1 $\times$ 3 $\times$ 4	2	55,424.	27,712.	2.11
Error (b) including quad. interaction	26	342,029.	13,155.	—
B. Within groups	384	345,260.	899.	13.66***
Trials	8	73,434.	9,179.	—
T $\times$ 1	16	8,312.	520.	—
T $\times$ 2	8	3,079.	385.	—
T $\times$ 3	8	9,234.	1,154.	1.72
T $\times$ 4	8	3,862.	483.	—
T $\times$ 1 $\times$ 2	16	28,980.	1,811.	2.69**
T $\times$ 1 $\times$ 3	16	16,531.	1,033.	1.54
T $\times$ 1 $\times$ 4	16	13,541.	846.	1.26
T $\times$ 2 $\times$ 3	8	1,764.	220.	—
T $\times$ 2 $\times$ 4	8	4,863.	608.	—
T $\times$ 3 $\times$ 4	8	4,289.	536.	—
Error (w) including higher interactions	264	177,371.	672.	—

— F less than 1.

\* Significant at the .05 level.

\*\* Significant at the .01 level.

\*\*\* Significant at the .001 level.

influenced the amount of time spent in communication, the rate of change of information in general, did not. However, when increased time was available for communication (decreased load), relatively more of it was utilized for sending messages in the high "demand" structure (LA) than in the low "demand" structure (HA).

The means and F ratio for trials indicate that total talking time increased with practice. The increase was greater for the LA structure than the HA structure.

## DISCUSSION

As might be expected, the results cast some light on the issues which gave rise to the study, but, at the same time, generate additional questions. The analysis of error scores supports the general hypothesis that "demands"

TABLE 6  
SUMMARY OF ANALYSIS OF VARIANCE OF  
"TOTAL TALKING TIME"

Source of Variation	df	Sums of Squares	Mean Squares	F Ratio
Total	431	585,627.	—	—
A. Between groups	47	537,455.	11,435.	—
Load	2	255,420.	127,710.	87.89***
Structure	1	181,507.	181,507.	124.92***
Procedure	1	8,348.	8,348.	5.74*
Performance	1	1,846.	1,846.	1.27
1 × 2	2	39,580.	19,790.	13.62**
1 × 3	2	675.	338.	—
1 × 4	2	2,778.	1,389.	—
2 × 3	1	430.	430.	—
2 × 4	1	2,766.	2,766.	1.90
3 × 4	1	263.	263.	—
1 × 2 × 3	2	1,526.	763.	—
1 × 2 × 4	2	801.	400.	—
1 × 3 × 4	2	3,374.	1,678.	1.16
2 × 3 × 4	1	360.	360.	—
Error (b) including quad. interaction	26	37,781.	1,453.	—
B. Within groups	384	48,172.	125.	—
Trials	8	5,488.	686.	6.53**
T × 1	16	1,176.	74.	—
T × 2	8	6,072.	759.	7.22**
T × 3	8	396.	50.	—
T × 4	8	603.	75.	—
T × 1 × 2	16	1,722.	108.	1.03
T × 1 × 3	16	2,380.	149.	1.41
T × 1 × 4	16	762.	48.	—
T × 2 × 3	8	459.	57.	—
T × 2 × 4	8	385.	48.	—
T × 3 × 4	8	1,108.	138.	1.31
Error (w) including higher interactions	264	27,621.	105.	—

— F less than 1.

\* Significant at the .05 level.

\*\* Significant at the .01 level.

\*\*\* Significant at the .001 level.

imposed by either the group's structure or task will affect mean performance level as well as the rate of error reduction. Increasing the communication interdependency between group members and/or increasing the rate of information input, increases the average number of errors and depresses the rate of improvement with practice. In addition, over-all group performance appears to be differentially effected by the procedure used for routing information, depending upon both structure and load. A volunteering procedure is better than a soliciting system only for the intermediate levels of "demand" imposed by structure and input variations.

The chief influence of structure and task demands, however, appears to be in setting the range over which learning occurs. The form of the curves for all experimental condi-

tions is generally quite similar; the parameters appear different. For none of the conditions is there evidence of discontinuities in the trends. Since the task appears to afford opportunity for rapid improvement in performance if optimum operating procedures are adopted, the lack of discontinuities may indicate that "insightful" reorganization did not occur, even under relatively low demand conditions.

Such an inference is in accord with the results obtained in several communication net studies. The net employed in the present study is, in essence, an all-channel or totally connected net with varying degrees of channel usage imposed by structure variations. Christie *et al.* (4) found some evidence to support a quasi-discontinuous notion of group learning in many of the nets they investigated, but little evidence of such a phenomena in a totally connected network. Guetzkow and Simon (5) demonstrated that groups in all-channel nets take relatively longer to achieve a stable and efficient organizational procedure than groups in wheel nets. Presumably as a consequence, the all-channel groups produce more continuous acquisition curves, while the wheel groups show rapid improvement in performance in the early trials and very little subsequent improvement.

The Guetzkow and Simon findings suggest that discontinuities may occur in the learning curve under conditions which facilitate the emergence of "optimum" operating procedures. In their study, the restrictions in communication which reduced the number of open channels, created an asymmetric position, and decreased the number of necessary relays, accomplished this. In the present experiment, elimination of unnecessary channels (possible in the HA structure) or restrictions on the relative availability of channels could possibly assist the development of efficient procedures, and if these procedures emerged, discontinuities in the learning curve might result.<sup>4</sup>

In general, the analysis of communication measures does not reveal as clear-cut effects

<sup>4</sup> It is possible, of course, that real progress in uncovering the factors influencing group learning will depend upon more detailed analysis of the component processes involved. In the present study error scores are composed of contributions from several "process" sources (e.g., translating instructions, detecting instrument reading changes, transmitting information) as well as several individuals, and errors from these various sources may smooth upon summation.



for the experimental conditions as does the error data. Surprisingly, the rate of change of input (load) effects no significant changes in communication behavior, although a longer total time available for communication increases the number of messages transmitted and the total time spent in communicating. Since the amount of information to be transmitted per trial is equal for all the loads it would be of interest to know what function is served by the larger number of messages under decreased load. The most plausible hypothesis is that they are redundant messages used for "checking."

Structure variation, which is coordinate with the volume of information to be transmitted, produced differences in all three measured aspects of communication behavior. In the LA structure more messages are sent, their average length is greater, and the total time spent in communicating is longer than in the HA structure. The differences are to be expected in view of the larger volume of information which had to be transmitted in the LA structure; however, the relative differences (approximately  $1\frac{1}{2}:1$ ) are not proportional to the differences in the required volume of information to be transmitted imposed by the two structures (4:1). Thus, relative to the demands imposed by the structure, the volume of transmitted information is much greater in the HA structure than in the LA structure. Since in both structures the amount of transmitted material was greater than that required by the task, it is probable that the differences reflect a much larger proportion of redundant messages in the HA structure than in the LA structure.

The difference is similar to that obtained for task load and appears to indicate that as demands decrease groups utilize the "free" time for checking activities. In view of the performance differences between load and structure conditions, this suggests that the primary determinant of performance adequacy was the degree of redundancy permitted by task and structure demands. The assumption that "free" time was employed in checking activities may also account for the lack of development of more efficient operating procedures. If, under the continued press of task requirements, all available time is devoted to activities directly task-relevant, reassessment

of over-all procedure is unlikely. If true, the provision of a break in the task during which group members could freely discuss their operation may be the most general method to facilitate the development of more efficient operating procedures.

With respect to the changes in communication behavior over trials, it is seen that with increasing practice more messages are transmitted, their average length decreases, and the proportion of the trial interval used for communication increases. Apparently, groups learn to "pack" more communication into a trial interval, but the rate of learning is relatively uniform. The change in "packing" efficiency could reflect the operation of several factors: the adoption of a coding procedure, the elimination of irrelevant material from messages, or increased efficiency in performing the observation and control action functions allowing more attention to be directed to communication. Whatever the source, there is some evidence that the more effective groups "pack" more efficiently than the less effective groups (they transmit shorter messages and use a larger proportion of the trial interval for communication).

Two aspects of these results are perplexing. One of the consistent findings of the communication net studies has been a decrease in the number of messages transmitted with increasing practice. Yet, in this study, communication behavior changes have as their common consequence an increased flow of information, even though in early trials the number of messages is much larger than required by the task. Several existing differences between the net studies and the present experiment may possibly account for the discrepancy.

The net studies employ a self-paced task with information from multiple sources to be screened for a single decision. The use of message cards obviates the need for storage of the transmitted material on the part of the Ss. The S's problem, therefore, is essentially in selecting targets for messages, and as efficiency in selection increases, the number of messages required for solution should decrease. The present study uses a forced-paced task (the periodic input to the group in no way depends on antecedent decisions) with multiple information inputs and multiple required decisions.

The information is transmitted verbally and no permanent record of messages is provided. Since every decision is dependent on collated information, a given message must reach *S* when he is free to receive it and relatively in-phase with other required information. The choice of a recipient being a trivial problem, the primary difficulty appears to be one of "phasing."

Rationally, it would seem that several procedures could reduce errors due to phasing difficulties: (a) advantage could be taken of predictable features of the input and the group could phase its information transmission with the predictable order of change in the information categories; (b) the source of information could initially transmit information once and repeat when requested; (c) the source could transmit repeated messages, thus allowing a recipient to sample this semicontinuous flow at his discretion.

Procedure (a) should lead to decreased information flow which is not consistent with the present results or with results obtained in a previous study which employed higher degrees of input predictability (7). Procedure (b) should result initially in greater volume of transmitted information than theoretically required although the flow should decrease with practice. Since communication differences obtained between the volunteering and soliciting procedures indicate that a strict volunteering procedure is not maintained, groups apparently did employ a procedure like (b). It is unlikely, however, that (b) alone could account for the increased information flow. Groups probably also reacted to the phasing difficulties by utilizing a technique such as (c).

The second perplexing aspect of the communication results is the apparent independence of the rate of change of communication behavior and the demands imposed by structure and task load. It seems justifiable to assume that environmental demands would be coordinate with drive strength and, thus, the greater the demands the more rapid the learning. Possibly, due to some process of social facilitation, the motivation for laboratory groups is higher (or lower) and less variable than is typical with individual *Ss* and the influence of external factors is attenuated. At any rate, further research on the relationships between the rate of modification of important

instrumental behaviors and group motivation and environmental demands is indicated.

The apparent independence of the rate of change of communication behavior and demands raises another more specific problem. Since the rate of error reduction is a function of demands, we are left in the peculiar position of trying to account for different rates of error reduction in terms of nondifferential rates of acquisition of important part-skills. One parsimonious explanation is that a given level of improvement in communication behavior has different consequences for error reduction in the various conditions. As the amount of information to be transmitted per unit time decreases, a given increase in communication efficiency could result in a proportionally greater decrease in errors. However, a precise statement of the relationship between rate of error reduction, demands of the task, and rate of improvement in communication behavior is impossible in the present study, and possibly in most studies of group learning. Methodologically, the result highlights the difficulty of accounting for group learning curves, based on traditionally employed error or time criteria, in terms of important part-processes. Similar difficulties have been encountered by other investigators (4, 8).

In conclusion, it should be pointed out that the usual strictures against overgeneralization apply quite literally in the present study. The experimental situation investigated was considerably more restricted than most group laboratory situations and the task problem was more structured than problems typically faced by either "real" groups or other temporary groups. It is interesting to note however, that despite such restrictions, much of the "explanation" of the data hinges on retrospective analysis of task requirements. Until a framework for the analysis of group task is available, hypotheses concerning the relationships between group structure and process, task demands, and group performance are going to be difficult to formulate and even more difficult to test.

#### SUMMARY

The present study was an extension of earlier experiments on the relationships between communication structure and group performance in an information processing task. The focus was



on changes in error performance and certain aspects of communication behavior as a function of practice under several levels of "demand" imposed by the group's communication structure and the rate of change of information input to the group. The analysis of results indicated:

1. Mean error performance and learning rate were a function of both communication structure and input load with better performance and faster rates for low demand conditions.

2. There was no evidence of discontinuities in the learning curves for any of the conditions and by inference no evidence that "insightful" reorganization occurred, even under low demand conditions.

3. All three aspects of communication behavior measured—average number of messages transmitted, average length of messages, and proportion of the trial interval spent in communication—varied with communication structure, but only the last varied with input load. There were changes in all three aspects with practice: more messages were transmitted, their average length decreased, and a greater proportion of trial time was spent in communication. Groups appear to learn to "pack" more information in a fixed interval with practice. However, there were no differences between conditions in the rate of change of communication behavior.

Apparently, left to themselves under constant situational pressures, groups modify their behavior at a characteristic rate which is independent of the demands placed on them. However, depending on the nature of the task, a given rate of behavior modification may have different consequences for the rate of improvement in some performance criterion such as errors.

## REFERENCES

1. BAVELAS, A. A mathematical model for group structure. *Appl. Anthropol.*, 1948, 7, 16-30.
2. BAVELAS, A. Communication patterns in task oriented groups. *J. acoust. soc. Amer.*, 1950, 22, 725-730.
3. CHRISTIE, L. S. Organization and information handling in task groups. *Operations Res. Soc. of America*, 1954, 4, 188-196.
4. CHRISTIE, L. S., LUCE, R. D., & MACY, J., JR. Communication and learning in task-oriented groups. Technical Report No. 231, Res. Lab. of Electronics, MIT, May 1952.
5. GUETZKOW, H., & SIMON, H. A. Impact of certain communication nets upon organization and performance in task oriented groups. *Management Science*, 1955, 1 (3) (4), 233-250.
6. HEISE, G. A., & MILLER, G. A. Problem-solving by small groups using various communication nets. *J. abnorm. soc. Psychol.*, 1951, 46, 327-335.
7. LANZETTA, J. T., & ROBY, T. B. Effects of work-group structure and certain task variables on group performance. *J. abnorm. soc. Psychol.*, 1956, 53, 307-314.
8. LUCE, R. D., MACY, J., JR., CHRISTIE, L. S., & HAY, D. H. Information flow in task-oriented groups. Technical Report No. 264, Res. Lab. of Electronics, MIT, August, 1953.
9. ROBY, T. B., & LANZETTA, J. T. Work group structure, communication, and group performance. *Sociometry*, 1956, 19 (2), 105-113.
10. ROBY, T. B., & LANZETTA, J. T. An investigation of task performance as a function of certain aspects of work-group structure. Lackland Air Force Base, San Antonio, Texas: Air Force Personnel & Train. Res. Center, June, 1956. (Research Report AFPTRC-IN-56-74)
11. ROBY, T. B., & LANZETTA, J. T. A replication study of work-group structure and task performance. Lackland Air Force Base, San Antonio, Texas: Air Force Personnel & Train. Res. Center. (Research Report in press)
12. SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *J. abnorm. soc. Psychol.*, 1954, 49, 547-553.

Received February 25, 1957. Prior publication.

## CRITIQUE AND NOTES

### A FURTHER STUDY OF THE EFFECT OF STRESS ON PALMAR PRINTS

ROBERT H. DAVIS

*The RAND Corporation<sup>1</sup>*

IN a recent publication, Mowrer (3) has described a method for obtaining palmar prints. Presumably densitometer measurements of such prints vary directly with emotional upset and relaxation. The terms emotional upset and relaxation are obviously difficult to tie down, but in a recent study (1) palmar prints were taken during periods which were "temporally proximal" to what are generally agreed to be stress inducing situations, oral examinations for the doctorate. Since the study reported by Beam was designed primarily to investigate the effect of real-life stress on conditioning and learning, the results with respect to palmar sweat under stress and non-stress conditions were not reported directly. It was demonstrated, however, that stress could be directly related to trials to reach a criterion of mastery, and furthermore, that the greater the increments in palmar sweating, the greater the increments in trials.

The present study was designed to test directly the hypothesis that real-life stress can be related to the index of palmar sweat. All measurements were taken during a period of stress rather than during a temporally adjacent period as in the study of Beam. Briefly, the palmar prints of a group of 34 college students were taken during an exami-

<sup>1</sup>The study described was conducted while the author was a member of the faculty of Allegheny College, Meadville, Pa.

nation and contrasted with control prints of the same Ss which were taken either two weeks prior to the examination or two weeks after the examination. The method utilized for taking the palmar prints was that described by Mowrer (3). Quantitative values were assigned to the prints by averaging five Kodak densitometer readings, one on each of the four fingers and the thumb. Rater reliability was assessed, using the rank-correlation coefficient. A separate coefficient was calculated for the experimental and control prints. Their values were .81 and .83 respectively. A nonparametric test was applied to the data. When palmar prints taken under the two conditions were contrasted using the sign test (2), densitometer values of the examination prints exceeded those of the control prints in 27 out of 34 cases ( $p < .01$ ), lending further support to the hypothesis that palmar prints can be meaningfully related to conditions of real-life stress.

#### REFERENCES

1. BEAM, J. C. Serial learning and conditioning under real life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543-551.
2. DIXON, W. J., & MASSEY, F. J. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.
3. MOWRER, O. H. *Psychotherapy: Theory and research*. New York: Ronald, 1953.

Received July 16, 1956.

### SITUATIONAL DIFFERENCES IN LEADER BEHAVIOR<sup>1</sup>

A. PAUL HARE

*Harvard University*

TWO types of leaders have been contrasted in the child study literature since the early 1930's, one group-oriented, the other, self-oriented (1, 6, 10, 11). In the present research, systematic observations of boys of each type are

made on the school playground, where these patterns have been observed in the past, and in the more intimate, less competitive situation of the back yard.

<sup>1</sup>Paper read at the annual meeting of the American Sociological Society, September, 1955. The present research is part of a continuing study of the emotional development of children which has been undertaken over the last five years by members of the Harvard School of Public Health and the Human Relations Service of Wellesley. The author is indebted to Erich Linde-

mann, the project sponsor, for his continuing support, to Donald Klein, Elizabeth Lindemann, and Ann Ross for their collaboration in developing the research design, and to Barbara Buckstein for carrying out half the field observation. The research was completed while the author was a Public Health Research Fellow of the National Institute of Mental Health.



The definitions of group and self-oriented leaders are derived from three previous studies by Gruber (7), Lindemann and Ross (9), and Vaughan and Faber (12) who use the same population of children from which the present sample is drawn. The group-oriented leader can be described as one who leads on the basis of self-confidence, has a clear perception of the requirements of the situation, and can play a follower's role. The self-oriented leader is anxious, aggressive, and has a driving need to be in the center of the group at all times. When these general definitions are described in terms of Bales' categories (2), the group-oriented leader would be expected to be higher than the self-oriented leader in the positive and neutral acts of categories 1 through 9 and lower in the negative acts of categories 10, 11, and 12 (see Table 1).

Previous studies have not indicated the nature of the input or the average behavior of other group members directed toward the leader. However, Lindemann and Ross have observed that each leader type tends to attract followers of the same type. One would, therefore, expect the input and output profiles to be similar for each leader style.<sup>2</sup>

Since most of the normative studies of child behavior are based on observations made in the school setting, situational differences in behavior are not reported. Barker and Wright have noted, however, that "stable, non-psychological situations such as Sunday school, day school, etc. coerce the children to behave in relatively homogeneous ways regardless of the individual characteristics of the children" (4).

Although the observations of boys on the school playground in the present study are in part a validation of the clinical judgment of certain teachers and social scientists, the more important objective of these observations is to delineate in detail the behavior of boys in two different natural settings.

#### METHOD

Six group-oriented leaders and six self-oriented leaders were selected from the third grades of four New England public schools on the basis of a teacher's rating and ratings by two previous observers of the boys.<sup>3</sup> The teachers were first asked to identify the boys in their classes who were leaders on the playground, second, to indicate whether the boys tended to be group-oriented or self-oriented, and third, to rank the boys of each type. From the teacher's list, boys were selected who had also appeared to be group- or self-oriented in

<sup>2</sup> In previous research by the author (8), two leader styles were imposed by boy leaders leading other boys in discussion groups. Although the major distinction between the two types of leaders was the amount of participation in the discussion, the results indicate that boys with different leader styles can produce significant differences in the interaction of their peers.

<sup>3</sup> Elizabeth Lindemann and Ann Ross had observed these children in the school over a period of three years.

their behavior in the research of the previous investigators.

During the winter and spring of 1954 these twelve nine-year-old boys were observed on the playground during a free play period and in the neighborhood, using the set of categories for interaction process analysis developed by Bales (2). (See Table 1.) The field observation technique differed from that of Barker (3) in that (a) interaction was scored rather than taken down verbatim, and (b) individual activity apart from the group was not scored. An observer scored only one boy at a time, recording all of his acts directed to others and all acts directed to him by other group members. When possible, each of two observers made half of the observations on each subject in each situation. Although the interobserver reliability approached but did not meet the standard suggested by Bales (2) at the beginning of the research, it was well above the standard by the time the last boy was observed.<sup>4</sup>

The average observation period on the school playground was four sessions for a total of 47 minutes (149 acts) and in the neighborhood was four sessions for a total of 75 minutes (218 acts). Interaction with teachers, parents, or observers was not recorded; only the interaction with other children was included in the profiles. Nonverbal interaction called for by the rules of the game was also omitted.

Bales' categories were adapted for field use by removing the restriction that category 4, "Gives suggestion," implies autonomy for the other.

#### RESULTS

The average percentage of the acts in each category by leader type and situation given in Tables 1 and 2 is obtained by converting the individual percentages by the arc sine transformation, computing the means, and reconverting to percentages. In the analysis of the data, however, the arc sine values are used.

An analysis of variance for each category of acts for the *output* data (Table 1) (acts directed by the leader to other group members) indicates that the group-oriented leaders give significantly more suggestions than the self-oriented leaders.<sup>5</sup> They also show less tension, a difference which approaches significance.

No significant differences appear in the *input* (acts received by the leader from the other members) between the two leader types (Table 2).

<sup>4</sup> When samples of mutually observed acts were compared, the reliability measure computed at the beginning of the research was 5.40 which exceeds the limit of 5.3 for 6 degrees of freedom suggested by Bales. By the end of the research the statistic was reduced to 2.74 where the maximum suggested is 6.3 for 7 degrees of freedom.

<sup>5</sup> The differences considered significant in this research have the probability of  $< .05$  of occurring by chance. Probabilities of  $< .10$  are referred to as approaching the significance level. A "one-tail" test is used only when the direction of the differences is predicted.

TABLE 1  
MEAN PERCENTAGE OF INTERACTION OUTPUT FROM  
GROUP- AND SELF-ORIENTED BOY LEADERS  
FOR PLAYGROUND AND NEIGHBORHOOD\*

Interaction Category	Playground		Neighborhood	
	G.-O. Ldrs.	S.-O. Ldrs.	G.-O. Ldrs.	S.-O. Ldrs.
1. Shows solidarity	5.7†	4.3	2.8	4.3
2. Shows tension release	6.4	8.5	7.8	6.9
3. Shows agreement	3.9	2.1	4.3	3.8
4. Gives suggestion	23.7	15.9	20.6	11.5
5. Gives opinion	7.2	7.3	6.5	8.0
6. Gives information	17.9	19.2	31.3	32.2
7. Asks for information	1.5	1.3	6.9	10.8
8. Asks for opinion	0.0	0.0	0.6	0.1
9. Asks for suggestion	0.0	0.0	0.1	0.0
10. Shows disagreement	1.4	3.1	4.3	3.2
11. Shows tension	4.3	10.5	3.0	4.5
12. Shows antagonism	20.1	19.5	6.1	10.5

\* Average number of acts per boy is 149 on the playground and 218 in the neighborhood.

† The raw percentage is converted by the arc sine transformation for calculation and reconverted for the table. For this reason the sum of percentages for each column does not equal exactly 100.

TABLE 2  
MEAN PERCENTAGES OF INTERACTION INPUT TO  
GROUP- AND SELF-ORIENTED BOY LEADERS  
FOR PLAYGROUND AND NEIGHBORHOOD\*

Interaction Category	Playground		Neighborhood	
	G.-O. Ldrs.	S.-O. Ldrs.	G.-O. Ldrs.	S.-O. Ldrs.
1. Shows solidarity	8.4	8.5	5.4	3.8
2. Shows tension release	0.8	2.3	4.1	4.9
3. Shows agreement	7.8	7.8	6.3	3.4
4. Gives suggestion	12.4	10.0	16.0	14.1
5. Gives opinion	7.6	8.5	7.2	7.3
6. Gives information	8.7	14.7	32.4	32.3
7. Asks for information	2.7	2.1	10.6	8.6
8. Asks for opinion	0.0	0.0	1.1	0.2
9. Asks for suggestion	0.0	0.2	0.1	0.1
10. Shows disagreement	3.3	5.7	4.4	4.8
11. Shows tension	0.8	1.0	0.6	1.4
12. Shows antagonism	24.0	24.3	4.7	13.4

\* Average number of acts per boy is 55 on the playground and 142 in the neighborhood.

Some additional differences between the two leader styles appear in only one of the two situations. The self-oriented leaders give and receive more antagonism than the group-oriented leaders in the neighborhood, a difference which approaches significance for output (Table 1) and is significant for input (Table 2). The self-oriented leaders also receive significantly less agreement in the neighbor-

hood and receive more information on the playground (a difference which approaches significance) (Table 2).

More significant differences in output appear between the two situations for the twelve leaders (Table 1). Giving and asking for information are both greater in the neighborhood and showing antagonism is greater on the playground. Differences in giving suggestion, disagreement, and showing tension approach significance with more suggestions and tension on the playground and more disagreement in the neighborhood for both leader types.

Some of the same differences between situations also appear in the input data (Table 2). Significantly more information is given and asked for in the neighborhood and more antagonism is shown on the playground.

The consistency of the behavior of each leader is analyzed by comparing his rank order in each category (a) for output between situations, (b) for input between situations, (c) for output and input on the playground, and (d) for output and input in the neighborhood. Only 54 of the 80 correlations thus generated are positive and of these only 11 are statistically significant, i.e., greater than .78 (correlations for categories 8 and 9 were not computed). This reflects a general lack of consistency in the observed behavior, less than that reported for adult subjects by Borgatta and Bales (5).

## SUMMARY AND CONCLUSIONS

Six group-oriented and six self-oriented boy leaders are observed on the school playground and in the neighborhood, using Bales' categories for interaction process analysis. The clinical judgments of teachers and social scientists are not entirely validated since boys of both leader types show the same amount of aggression on the school playground, although the self-oriented leader is more aggressive in the neighborhood. Other expected differences appear since the self-oriented leader gives fewer suggestions and shows more tension in both situations.

It is possible that the group-oriented leaders are perceived as less aggressive on the school playground because they have a more legitimate reason for aggression in the interest of group control. The self-oriented leader uses his aggression to maintain his central position so that he is usually described as "more aggressive," although both types have the same number of aggressive acts on the school playground.

The major finding of this research is that there are more differences between situations than between the two leader styles. For the twelve nine-year-old boy leaders in this sample, giving and asking for information and disagreement are higher



in the neighborhood, and giving suggestions, showing tension, and showing antagonism are higher on the playground. Differences in input are in the same direction, with more information given and asked for in the neighborhood, and more antagonism shown on the playground.

The higher amount of information in the neighborhood probably results from the fact that the children spend more time making up games, while the increase of suggestion, tension, and antagonism on the playground reflects a more competitive situation in which the individual must assert himself to maintain his status in a large group composed of children his own age.

## REFERENCES

1. ANDERSON, H. H. Domination and socially integrative behavior. In R. G. Barker, J. S. Kounin, & H. F. Wright (Eds.), *Child behavior and development*. New York: McGraw-Hill, 1943. Pp. 459-483.
2. BALES, R. F. *Interaction process analysis*. Cambridge, Mass.: Addison-Wesley, 1950.
3. BARKER, R. G. *One boy's day*. New York: Harper, 1951.
4. BARKER, R. G., & WRIGHT, H. F. Psychological ecology and the problem of psychosocial development. *Child Developm.*, 1949, 20 (3), 131-143.
5. BORGATTA, E. F., & BALES, R. F. The consistency

of subject behavior and the reliability of scoring in interaction process analysis. *Amer. sociol. Rev.*, 1953, 18, 566-569.

6. BUHLER, CHARLOTTE. Social behavior of the child. In C. A. Murchison (Ed.), *A handbook of child psychology*. Worcester, Mass.: Clark Univer. Press, 1931.
7. GRUBER, S. The concept of task orientation in the analysis of play behavior of children entering kindergarten. *Amer. J. Orthopsychiat.*, 1954, 24, 326-335.
8. HARE, A. P. Small group discussions with participatory and supervisory leadership. *J. abnorm. soc. Psychol.*, 1953, 48, 273-275.
9. LINDEMANN, ELIZABETH B., & ROSS, ANN. A follow-up study of a predictive test of social adaptation in preschool children. In G. Caplan (Ed.), *Emotional problems of early childhood*. New York: Basic Books, 1955, Pp. 79-93.
10. PARTEN, MILDRED B. Leadership among preschool children. *J. abnorm. soc. Psychol.*, 1933, 27, 430-440.
11. FIGORS, P. *Leadership or domination*. Boston: Houghton-Mifflin, 1935.
12. VAUGHN, W. T. JR., & FABER, E. Field methods and techniques: The systematic observation of kindergarten children. *Hum. Organization*, 1952, 11, 33-36.

Received July 17, 1956.

RELIGIOUS PRACTICE AND CARDIOVASCULAR REACTIONS DURING STRESS<sup>1</sup>

STANLEY H. KING

AND

DANIEL H. FUNKENSTEIN

Graduate School of Public Health,  
University of Pittsburgh

Harvard Medical School

RECENT research has pointed to a number of psychological and sociological correlates of physiological reactions during acute stress. In terms of cardiovascular response we have demonstrated a positive association between anger directed outward and a nor-epinephrine-like cardiovascular pattern and between anger directed inward or anxiety and an epinephrine-like pattern (3). We have also shown that a subject's perception of his father as stern and dominant in discipline was associated with a nor-epinephrine-like cardiovascular reaction while perception of father as mild and nondominant in discipline was

associated with an epinephrine-like pattern (6). This report describes a third correlate of cardiovascular reactions in acute stress, that of religious practice and attitudes.

## METHOD

The data on which these findings are based were gathered in a two-year study of healthy college students during acute laboratory stress. The details of this study have been reported elsewhere (2, 3) but in brief the procedure was as follows. One hundred twenty-five randomly selected male students were placed in laboratory stress situations where they were given arithmetical problems to do in their heads while being hurried and criticized by the experimenter, or they were asked to repeat a story from memory while talking into an auditory feedback device. Cardiovascular measurements by means of a blood pressure cuff and a ballistocardiograph were obtained before and after each session and the Ss were interviewed at the end of the stress situation for a measure of their emotional reactions. In addition, each S was seen by the Es on other occasions

<sup>1</sup> Grateful acknowledgement is given for support of this work to the School of Aviation Medicine, Randolph Field, Texas, under Contract AF 33(038)20142, The Russell Sage Foundation, the Supreme Council, 33rd Degree, Scottish Rite, Northern Masonic Jurisdiction, U. S. A., through the National Association for Mental Health, and the Laboratory of Social Relations, Harvard University.

where a battery of psychological tests and a biographical questionnaire were administered.

On the basis of relative changes between the pre-stress and stress measures of blood pressure and ballistocardiograph tracings the Ss could be divided into two main groups comparable to those described earlier by Funkenstein and Meade (4). In their study blood pressure and ballistocardiograph tracings were taken from normal Ss during the infusion of nor-epinephrine and during the infusion of epinephrine. The various physiological measures fell into two distinct patterns under each condition, characterized by changes in pulse, mean IJ wave on the ballistocardiograph, and mean blood pressure. The exact procedure for the determination of these two patterns can be found elsewhere (4, 5). The pattern of physiological variables in one group of Ss in the present study was like that found by Funkenstein and Meade during the infusion of nor-epinephrine and is thus termed *nor-epinephrine-like*. The other group had a pattern of variables comparable to that found during the infusion of epinephrine, and we have termed it *epinephrine-like*.

Data on religious practice and attitudes were obtained from the personal history questionnaire and the Religious Conventionalism scale of Levinson.

On the questionnaire each S was asked about the churchgoing behavior of his parents through the following questions:

1. How frequently does your father attend church?  
a. Frequently; b. occasionally; c. seldom.
2. How frequently does your mother attend church?  
a. Frequently; b. occasionally; c. seldom.

Responses to these questions were combined into a single measure of parental church attendance, as follows:

1. Regular attendance (both attend frequently or one attends frequently and the other occasionally);
2. occasional attendance (both attend occasionally or one attends frequently and the other seldom);
3. infrequent attendance (both attend seldom or one attends seldom and the other occasionally).

Further data in the area of religion were available from the scores on the Religious Conventionalism scale which had been administered as part of the battery of attitude scales developed by Levinson out of the work described in *The Authoritarian Personality* (1). Levinson has described this particular scale as follows (7):

The *Religious Conventionalism* (RC) scale contains a variety of ideas about the church, the Bible, God, prayer, and the like. High scores reflect a conception of God as punishing power figure and of the church as absolute moral authority, as well as a marked emphasis on faith, tradition, and conformity to institutional forms. Low scores, on the other hand, represent a religious humanistic or a nontheistic approach which emphasizes reason, personally derived values, and a naturalistic rather than supernaturalistic view of the world.

The S is presented with 12 statements in the area of religious conventionalism and asked to indicate the degree of agreement or disagreement with each statement on a seven-point scale ranging from strongly agree to strongly disagree. He can receive a score of 1-7 on each statement or a range of 12-84 on the total.

TABLE 1  
PARENTAL CHURCH ATTENDANCE AND CARDIOVASCULAR REACTIONS OF SUBJECTS DURING STRESS

Parental Church Attendance	Cardiovascular Pattern of Subject		Total
	Nor-epinephrine-like	Epinephrine-like	
Regular	20	11	31
Occasional	13	19	32
Infrequent	3	12	15
Total	36	42	78

$$\chi^2 = 8.719.$$

$$p < .02.$$

TABLE 2  
SCORES OF THE CARDIOVASCULAR REACTION GROUPS ON THE RELIGIOUS CONVENTIONALISM SCALE

Cardiovascular Reaction Group	Religious Conventionalism Scale		
	N	Mean	SD
Nor-epinephrine-like	36	42.94	19.00
Epinephrine-like	41*	33.63	18.79

\* One subject became unavailable for testing after filling out the personal history questionnaire and before taking the RC scale.

$$t = 2.13.$$

$$p < .05.$$

## RESULTS

The association between the measure of parental church attendance and the cardiovascular reactions of the Ss in stress can be seen in Table 1. The Ss whose parents were regular attenders were more inclined to have *nor-epinephrine-like* cardiovascular reactions while those whose parents were occasional or infrequent attenders had *epinephrine-like* patterns. This difference was especially striking between Ss with regularly attending parents and those with infrequently attending parents. The chi square in this table was 8.719 which with 2 degrees of freedom is significant at the .02 level of confidence.

The mean RC scale scores for the *nor-epinephrine-like* and *epinephrine-like* groups are shown in Table 2. It will be noted that even though the SD was large there was still a significant difference between the means with the use of a *t* test. Those with *N* patterns had higher RC scale scores than those with *E* patterns.

## DISCUSSION

These data when combined with those on which we have reported previously indicate that there is a constellation of psychological and sociological factors which are associated with the cardiovascular reactions of healthy Ss in acute stress. This constellation includes the S's immediate emotional



reaction, his attitudes in the area of religious values, his perception of parental behavior in discipline, and the churchgoing behavior of his parents. The Ss who responded to acute stress with a nor-epinephrine-like cardiovascular reaction tended to also respond with anger directed at the E. In addition, they were individuals who perceived their father as the principal disciplinarian in the family and stern in his discipline, they had conservative attitudes in the area of religious faith and belief, and they came from families where the church attendance habits were regular. On the other hand, those Ss who responded to acute stress with an epinephrine-like reaction tended to respond with anger directed against themselves or anxiety. In addition, they perceived their fathers as non-dominant in discipline, had moderate to liberal attitudes in the area of religious faith and belief, and came from families where the pattern of church attendance was occasional or infrequent.

These factors come from a number of levels, not only the immediate biological (physiological and emotional), but the more enduring organization of perceptions (perception of father and attitudes toward religion), and even the surrounding family milieu (church attendance of parents).

We leave it to future research to spell out the manifold implications of these associations. We do suggest that the relationships are of sufficient

strength to encourage further interdisciplinary research among the fields of physiology, psychology and sociology.

## REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper & Bros., 1950.
2. FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. The experimental evocation of stress. In *Symposium on Stress*, Washington, D.C., Walter Reed Army Medical Center and Army Medical Service Graduate School, 1953, pp. 304-322.
3. FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. The direction of anger during a laboratory stress-inducing situation. *Psychosom. Med.*, 1954, 16(5), 404-413.
4. FUNKENSTEIN, D. H., & MEADE, L. Nor-epinephrine-like and epinephrine-like substances and the elevation of blood pressure during acute stress. *J. nerv. ment. Dis.*, 1954, 119(5), 380-397.
5. KING, S. H. Emotional and cardiovascular responses during stress. Unpublished doctor's dissertation, Harvard Univer., 1953.
6. KING, S. H., & HENRY, A. F. Aggression and cardiovascular reactions related to parental control over behavior. *J. abnorm. soc. Psychol.*, 1955, 50, 206-210.
7. LEVINSON, D. J. The intergroup relations workshop: Its psychological aims and effects. *J. Psychol.*, 1954, 38, 103-126.

Received June 26, 1956.

CERTAINTY OF JUDGMENT AND RESISTANCE TO SOCIAL INFLUENCE<sup>1</sup>

HAROLD H. KELLEY

*University of Minnesota*

AND THOMAS W. LAMB

*Harvard Medical School*

SEVERAL studies have indicated that social influences have greater effect on judgments the more ambiguous is the stimulus being judged (2, 6). One explanation of this finding is that under conditions of ambiguity, individuals are less certain about their judgments and, hence, more easily influenced. The factor of certainty has been studied independently of stimulus ambiguity by Hochbaum (4) who found high confidence in an opinion to be associated with high resistance to influence. The present note presents further evi-

dence on the importance of certainty, obtained under conditions where there are marked individual differences in reactions to a given stimulus and associated differences in certainty of judgment.

The method involved the use of phenylthiourea (also known as phenylthiocarbamide), hereafter referred to as PTU. This substance is tasteless to some individuals but tastes extremely bitter to others. After preliminary work indicated that we could present PTU in a fairly controlled manner and readily distinguish tasters from nontasters,<sup>2</sup>

<sup>1</sup> Conducted while the authors were at Yale University, under the auspices of the Yale Communication Research program. The authors are indebted to several chemists at Yale and Minnesota for their advice.

<sup>2</sup> Blakeslee (1) has found that tasting behaves as a dominant gene and nontasting as a recessive gene. The basis for this difference is now thought to exist in the

we used it to study the following question: Does the effect of a majority depend upon the particular opinion they advocate, relative to the opinion held by the minority? For example, will the effect of the majority's espousal of "A" when the minority initially favors "B" be the same as if the majority had supported "B" and the minority, "A"? That this problem readily lends itself to the use of PTU is apparent: We have merely to form groups composed of unequal numbers of tasters and nontasters, in some of which the tasters are in the majority and in others of which the nontasters predominate.

### METHOD

A rating form was prepared with three scales of "pleasantness of taste," ranging from 0 (bad) to 10 (good). The scales were labeled A, B, and C, respectively, to correspond to three gummed labels, also labeled A, B, and C, which were attached to the bottom of the forms. Label A was a plain label, and Label B

interference of PTU with the activity of certain enzymes (7, Volume I, p. 337). It is also thought to be a goiterogenic compound if administered on a long-term basis, as it inhibits the formation of thyroxine, a hormone of the thyroid gland (7, Volume II, p. 427). It has been used as a food preservative, but recently in this connection the question of its possible carcinogenic action has been raised (5).

PTU can be obtained in a pure form from the Eastman Kodak Company or in a semipure form from many chemical supply houses. It can be recrystallized from alcohol, since it is extremely soluble in hot alcohol and relatively insoluble in cold alcohol. A negative cyanide test should be obtained before the chemical is used on Ss (3, p. 258).

Because of the possible dangers of large quantities of the compound and the nauseating effects on some persons, it is desirable to present PTU in very small quantities. At the same time, trial taste-tests indicated that strong concentrations should be used if an immediate, relatively standard reaction is desired. With weak concentrations many tasters have a sensation of bitterness only as an "aftertaste" which occurs up to ten seconds after the presentation. These considerations were met by presenting PTU in the form of a dried solution on the back of ordinary gummed labels. This method provides the S with a high concentration of PTU, but at the same time insures that he ingests only a minute quantity of it.

The solution was prepared by diluting to twice its volume a saturated solution of PTU in Baker Extract Company's "True Fruit Cherry Flavor" which has a 16% alcohol base. It was then applied with an atomizer to the back of Reyburn's label No. 13 until a deep purple color was attained. This rather crude method of "dosage" was found to produce quite standard results, because of the high concentration of PTU that was used. Individual differences in reaction were noted, but of the Ss to whom we presented the labels, not one exhibited a violent reaction, yet all the "tasters" experienced an intense, bitter taste.

was the same kind of label coated with pineapple flavor. Label C was prepared with a combined coating of PTU and cherry flavoring. Scales B and C each had an accompanying scale for degree of certainty, ranging from 1 (not at all certain) to 7 (very certain).

The subjects (Ss), Yale University upperclassmen, were first seen individually and privately. Each was told that we were trying to find out "whether there are absolute values for these tastes or whether there are individual differences in ratings of them." He was then given a form, asked to taste Label A, and instructed to give it a rating of 5 (average) on the scale of pleasantness. He was then asked to taste B and C, and to rate each one in relation to A. After rating B, he was asked to rate his certainty that B should be given the pleasantness rating he had given it. A similar procedure was followed for C.

On the basis of these private judgments, the Ss were scheduled in 30 three-man groups, half composed of one taster and two nontasters and half, of two tasters and one nontaster. Because the population of Ss yielded a ratio of tasters to nontasters of approximately 3:1, about 150 individuals were privately tested before the 30 groups could be arranged. In scheduling each group, the further requirement was made that each of the three men should know the other two, the result being that men who lived near each other in a given residential college were placed in the same group.

The groups were assembled one to three days after the individual testing. The instructions were similar to those used earlier, except that certainty ratings were not requested and the Ss were asked to announce their ratings publicly, each one first giving his rating of B, and then of C. Each time, the majority persons, whether tasters or nontasters, were asked to announce their ratings first. When it came his turn on B, the minority person usually found that the others had given ratings similar to his private ones. But on C, he found that their ratings were markedly different from his own evaluation.

### RESULTS

The data were analyzed only for the 90 Ss used in the group portion of the experiment.

1. The tasters and nontasters agreed very closely in their ratings of Label B, the one coated with pineapple flavoring. On the average, they gave it a rating near 7.0 on both the initial private ratings and the publicly announced ones. We mention this to show that the tasters and nontasters are comparable in their reactions to a substance not containing PTU and that they used the rating scale in much the same manner.

2. The tasters' initial private ratings of C, the experimental label coated with PTU and cherry flavoring, were markedly lower than the ratings made by the nontasters. The mean rating for the 45 tasters was 0.8 and for the 45 nontasters, 6.7. A *t* test of this difference yields a value of 32.7 which is obviously significant. The two distributions of ratings do not overlap at all. The range of ratings given by the tasters is from 0 to 2; for the



nontasters, from 4 to 8, with 42 of the 45 giving ratings from 6 to 8.

3. The public judgments of Label C were first examined for the Ss in each of the 30 groups who announced their ratings first. These ratings, when compared with the private ones made by the same 30 persons several days earlier, indicate the stability of these taste judgments in the absence of experienced disagreement. The scatterplot of private versus public (early vs. late) ratings reveals two separate clusters of Ss, the tasters and the nontasters. The product-moment correlation between the early and late ratings is  $+0.98$ . Put in other terms, the data showed that 19 of the 30 Ss gave identical ratings on both occasions, 10 gave ratings which differ by one point, and 1 gave ratings differing by two points. The results were virtually identical for the Ss (also members of the majorities) who announced their ratings second. It is quite clear, then, that in the absence of pressure to change, the private ratings were highly predictive of the later public ones.

4. The public ratings of C were next examined with respect to the changes from the earlier private ratings exhibited by the Ss who announced their judgments last. Since these Ss were always in the minority, these changes indicate the effects of the majority opinions. There is a marked difference between the minorities composed of tasters and those composed of nontasters. The average amount of change for the tasters was 0.0, whereas the average for the nontasters was  $-2.6$ , the negative sign indicating that they made a lower (less "pleasant") rating in the group situation. There is clearly a significant difference between the two kinds of Ss in their tendency to go along with the majority opinion. Only 4 of the 15 tasters shifted in the indicated direction (toward a higher rating) while 13 of the 15 nontasters shifted toward a lower rating (with Yates' correction, chi square =  $8.7$ ,  $p = .004$ ). It appears, then, that a taster majority has a greater effect on a nontaster minority than a nontaster majority has on a taster minority.

Two possible bases for this difference are suggested by our data, only the second of which seems to be important. First, it happened that the average discrepancy between the ratings announced by the majority members and the private rating previously given by the minority member was somewhat larger in the groups where the tasters were in the minority. While this difference between the two kinds of groups approaches significance ( $p = \text{approx. } .06$ ), we doubt that it accounts for the observed effect. It seems unlikely that the minority tasters would have changed less simply because the majority members' opinions were somewhat farther away from their private ones. In fact, within each sample, there seems to be the opposite tendency: greater change tends

to be associated with a larger discrepancy. If we compare Ss for whom the discrepancy was identical in size, there is still a marked tendency for nontasters to exhibit more change than tasters.

Second, tasters and nontasters were found to differ in how certain they were about their ratings of the pleasantness of Label C. The average certainty rating for the tasters was significantly higher than that for the nontasters ( $6.6$  as compared with  $5.4$ ;  $t = 6.25$ ,  $p < .001$ ). This difference was also significant, at the .01 level, when only the Ss used as minority members were compared. This result suggests that the reason for the tasters' great resistance to majority influence may be found in the greater intensity of their opinions about PTU. The correlation between change and certainty within each sample is equivocal on this point because the range of both variables is severely restricted when we consider only the tasters or nontasters. We may note parenthetically that for both Labels B and C, certainty shows the typical U-shaped relation to extremity of rating on pleasantness. The closer a person rated the flavor to the extreme ends of the scale, the more certain he was of the correctness of his rating. The eta coefficient is highly significant in both cases.

#### SUMMARY

In a simple social influence problem involving judgments of taste with PTU the critical stimulus, it was found that the effects of majority opinion were markedly different depending upon whether the minority persons were tasters or nontasters of PTU. The data suggest that this asymmetric effect, whereby nontasters are more susceptible to majority influence, may be attributable to the stronger reactions tasters have to PTU and the resulting greater certainty they have about their judgments of the substance.

#### REFERENCES

1. BLAKESLEE, A. L. Genetics of sensory thresholds: Taste for phenylthiocarbamide. *Proc. nat. Acad. Sci.*, 1932, **18**, 120-130.
2. COFFIN, T. E. Some conditions of suggestion and suggestibility. *Psychol. Monogr.*, 1941, **53**, No. 4.
3. FEIGL, F. *Spot tests. Vol. I. Inorganic applications.* Houston: Elsevier, 1954.
4. HOCHBAUM, G. M. The relation between group members' self-confidence and their reactions to group pressures to uniformity. *Amer. sociol. Rev.*, 1954, **19**, 678-687.
5. LEAR, J. Food and cancer. *Saturday Rev.*, 1956, **39**, No. 40, 57-62.
6. LUCHINS, A. S. Social influences on perception of complex drawings. *J. soc. Psychol.*, 1945, **21**, 257-263.
7. SUMNER, J. B., & MYRBACK, K. (Eds.), *The enzymes.* New York: Academic Press, 1950-1952.

Received August 10, 1956.

## TWO KINDS OF ASSUMED SIMILARITY BETWEEN OPPOSITES<sup>1</sup>

IVAN D. STEINER AND COLIN G. McDIARMID

*University of Illinois*

IN a series of recent studies, F. E. Fiedler (3, 4, 5) has shown that a leader's "assumed similarity between opposites" (ASo) is an important correlate of his team's efficiency. Basketball teams and surveying teams were shown to be more efficient when the sociometrically most-chosen member attributed very different psychological traits to persons he regarded as good and bad co-workers. The efficiency of bomber and tank crews was also found to be related to the leader's ASo. This paper reports an investigation of some statistical and psychological properties of ASo.

ASo is a measure of the magnitude of the differences between two profiles. One profile represents the individual's ratings of a good co-worker on twenty or more six-step graphic rating scales. The ends of each scale are identified by polar terms such as "energetic" and "tired." The second profile consists of the individual's ratings of a poor co-worker on the same graphic rating scales. The ASo score is obtained by summing the squares of the distances separating these two profiles from one another (i.e., the distances separating the ratings given to the good co-worker from the ratings given to the bad co-worker). Thus:

$$ASo = \sum_{i=1}^N (g - b)^2$$

when  $(g - b)$  equals the discrepancy between the ratings given to the good and the bad co-workers on one of the graphic rating scales, and  $N$  equals the number of rating scales employed. The ASo score may be broken into two component parts (2):

$$\frac{ASo}{N} = (M_g - M_b)^2 + \sigma_{g-b}^2$$

where  $M$  is the mean of the several ratings given to a co-worker. The  $(M_g - M_b)^2$  component is the square of the average *elevation difference* between the two profiles—the tendency of one profile to lie above the other. Intuitively, we can view this component as reflecting the individual's tendency to regard one co-worker as endowed with more desirable attributes than the other. The second component measures the degree to which the two profiles have different shapes, regardless of their average elevations. This component may be viewed as reflecting the individual's tendency to regard his two co-workers as possessing *different combinations* of good and bad traits.

<sup>1</sup> This research was partially supported by a grant from the University of Illinois Research Board.

There is no logical necessity for these two components of the ASo score to be correlated. It is entirely possible for a small or zero average elevation difference to occur along with a large difference in profile shapes. And a large average elevation difference could occur with a small difference in profile shapes. The ASo score is, therefore, a gross measure which combines two logically independent measures. The purpose of this research is to determine whether the two components of ASo are also empirically independent of one another, and, if so, to determine whether the two components correlate differently with other measures. The other measures employed in this research include scores on the California F scale (1), and scores obtained from self-ratings. Reasons for expecting scores on the F scale to correlate with one or both components of ASo will be discussed later.

### METHOD

Forty-two members of an undergraduate class in psychology served as subjects (Ss) for this study, but because of omissions from the data, the  $N$ s involved in the reported results range from 38 to 42. The two sexes were approximately equally represented in the usable samples.

#### ASo Scores

Ss responded to Fiedler's instrument consisting of 20 graphic rating scales, the ends of which were identified by polar terms. As is customary with this instrument, Ss rated themselves, a good co-worker, and a bad co-worker on each of the 20 scales. The Ss were asked to think of someone with whom they had been able to work very well on some occasion in the past, and of someone with whom they had had difficulty working. Each S's ASo score was computed by obtaining the sum of the squared distances separating his ratings of these two co-workers.

#### Elevation Difference Scores

In order to compute the  $(M_g - M_b)$  component of ASo it was necessary to obtain judgments concerning which end of each scale was regarded by S as the good end. For this purpose each S rated himself on Fiedler's instrument two additional times, once giving himself the benefit of any doubts whatever, and once without giving himself the benefit of any doubts. Inspection of these two sets of self-ratings permitted a determination of the good-bad direction of the scale for the S. Ratings at the extreme bad end of a scale were then assigned a value of one and those at the good end were assigned a value of 6. The mean value of the ratings given to the good co-worker was obtained, as was the mean value of ratings given to the bad co-worker. The difference between these two means is the S's Elevation Difference score.



### Profile Dissimilarity Scores

Each *S*'s Profile Dissimilarity score was obtained by subtracting his squared Elevation Difference score from his  $\frac{ASo}{N}$  score. The Profile Dissimilarity score is the square root of the difference obtained in this manner.

### Self-Ratings

As noted above, each *S* rated himself three times: once attempting to be as accurate as possible (realistic self-ratings), once giving himself the benefit of any doubts (optimistic self-ratings), and once without giving himself the benefit of any doubts (pessimistic self-ratings). The mean and standard deviation of these three sets of self-ratings were computed for each *S*.

*F Scale.* A month after the previous data had been obtained each *S* responded to the *F* scale. Items were scored in the standard manner.<sup>2</sup>

### RESULTS

As should be expected, both Elevation Difference scores and Profile Dissimilarity scores were found to be positively correlated with ASo scores. The product-moment correlations are .75 and .53 respectively. With a sample of 42 persons, both correlations are significantly different from zero at the .01 level. The correlation between Elevation Difference scores and Profile Dissimilarity scores is -.03. The two components of ASo appear to be empirically independent of one another.

Relationships between each of the three measures and other variables are indicated by the product-moment correlations shown in Table 1. Three different patterns of relationships are discernible. *F* scores correlate significantly with ASo scores, but correlations with the two components of ASo are low and do not reach statistical significance. Standard deviations of self-ratings do not correlate significantly with either ASo scores or Elevation Difference scores, but do correlate significantly with Profile Dissimilarity scores. Finally, the mean of the individual's optimistic self-ratings (i.e., the mean of the self-ratings made by *S* when he was instructed to give himself the benefit of any doubts) fails to correlate significantly with either ASo scores or Profile Dissimilarity scores, but does correlate significantly with Elevation Difference scores. (Neither the mean "realistic" self-rating scores nor the mean "pessimistic" self-rating scores correlated significantly with ASo or either of its components.)

These findings thus indicate that the two components of ASo sometimes correlate in different degrees (and possibly in different directions) with

TABLE 1  
CORRELATION OF ASo AND ITS TWO COMPONENTS TO OTHER VARIABLES

	F Scale Scores	Mean of Self-ratings	Sigma of Self-ratings	Mean of Optimistic Self-ratings
	N			
	38	42	42	42
ASo scores	-.43**	.02	.29	.21
Elevation Difference Scores	-.21	.18	-.07	.35*
Profile Dissimilarity Scores	-.29	-.09	.62**	-.10

\* Reaches the .05 level.

\*\* Reaches the .01 level.

other variables. The components may also combine to produce ASo scores which correlate more highly with other variables than do either of the components taken separately.

### DISCUSSION

The ASo score, according to these results, is a gross measure that combines two uncorrelated variables. One of these variables is the perceived discrepancy between the overall "goodness" of the traits possessed by two co-workers; the other variable is the perceived dissimilarity between the patterns of traits possessed by two co-workers. Each of these two variables contributes substantially to the ASo score, but two *S*s may obtain the same ASo score even though their scores on the two components of ASo are very different. For example, two *S*s in the present study obtained ASo scores of 166; for one *S* the  $(M_o - M_b)^2$  component was responsible for seven per cent of the ASo score, whereas for the other *S* this component was responsible for 82 per cent. For most purposes it would probably be unrealistic to treat these two persons as though their ASo scores were equivalent.

When ASo scores are related to an external variable, a significant correlation may be obtained because either or both of the components of ASo are correlated with the external variable. Or, an insignificant correlation may be obtained because neither component is correlated with the external variable, or because both are correlated but in opposite directions. Since it is quite possible for one or both components to correlate much more highly with an external variable than does ASo itself, it is recommended that studies of "assumed similarity between opposites" deal with the components as well as with ASo.

In this research the direction of the relationships between *F* scale scores and the two components of ASo was not predicted in advance, but significant relationships were expected. According

<sup>2</sup> Form 45 of the *F* scale (1, pp. 255-257) was employed in this research, but item number 22 was deleted because it had become obsolete since the time the scale was published.

to Adorno et al. (1), the high F scorer is reluctant to see differences between members of his in-groups; all members are good, whereas people in out-groups are bad. On the other hand, the low F scorer is described as willing to see various combinations of traits in members of his in-group, and to minimize differences between members of in-groups and out-groups. Consequently, if one assumes that both the good and the bad co-workers are regarded by the S as members of his in-group, both components of ASo should be small for high F scorers and large for low F scorers. But, if the poor co-worker is regarded as a member of an out-group, both components of ASo should be larger for the high F scorer than for the low F scorer. The findings reported in Table 1 suggest that when people are asked to think of someone with whom they have worked poorly, they tend to think of an in-group member.

There are two possible interpretations of the .62 correlation between Profile Dissimilarity scores and the standard deviations of Ss' realistic self-ratings. Perhaps a response set leads some people to give ratings that are highly variable from item to item, and leads other people to give ratings that are rather constant from item to item. If these response sets operated both when S rated himself and when he rated his good and bad co-workers, a correlation between Profile Dissimilarity scores and the standard deviation of realistic self-ratings might result from the response set alone. However, if such response sets are operating, they should be expected to affect Ss' optimistic and pessimistic self-ratings just as they affect ratings of good and bad co-workers. Inspection of the data shows that when standard deviations of Ss' optimistic and pessimistic self-ratings are computed and added to the standard deviations of their realistic self-ratings, the correlation between the sums of these three standard deviations and the Profile Dissimilarity scores falls to .20. This finding tends to contradict an explanation based on the assumption of a broad, "mechanical" response set, and suggests a second explanation. Possibly some persons really do see themselves and other people as the possessors of a mixture of good and bad traits, whereas other persons tend to see both themselves and their associates as the possessors of rather uniformly good or uniformly bad traits. The data of this research do not provide a firm basis for choice between these two explanations. However, whatever the proper explanation may be, it is clear that Profile Dissimilarity scores represent a variable and potentially significant component of ASo, and that this component is empirically independent of the Elevation Difference component.

The findings of this study suggest that Fiedler's highly productive research might be made even more valuable if ASo scores were broken into their component parts. Such a step might also be helpful in clarifying the reasons why the ASo of a leader is related to his group's efficiency. Fiedler's explanation of his results rests upon assumptions concerning the personalities of high and low ASo leaders. For example, the leader with a large ASo score is said to be "critical, reserved, and rejecting, and especially so toward men who he considers to be poor co-workers." He is presumed to discriminate sharply between good and bad co-workers, to resist emotional involvement, and to be guided by task-oriented attitudes. These assumptions concerning the personal qualities of leaders with large ASo scores have not yet been tested, and Fiedler himself has indicated that he regards them as tentative. In light of the present study it is reasonable to guess that leaders with identical ASo scores may have quite different personal qualities depending upon the size of the components of their ASo scores. Somewhat more complex assumptions concerning the relationships between ASo and personality variables would seem to be needed.

#### SUMMARY

The conventional measure of assumed similarity between opposites (ASo) can be broken into two components which are logically and empirically independent of one another. Taken separately these two components sometimes correlate more highly with external variables than does ASo itself. Implications of these qualities of ASo are discussed.

#### REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper & Bros., 1950.
2. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **62**, 177-193.
3. FIEDLER, F. E. *Assumed similarity measures as predictors of team effectiveness in surveying*. Urbana, Ill.: Bureau of Research & Service, Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 6, Contract N6ori-07135.)
4. FIEDLER, F. E. The influence of leader-keyman relations on combat crew effectiveness. *J. abnorm. soc. Psychol.*, 1955, **51**, 227-235.
5. FIEDLER, F. E., HARTMANN, W., & RUDIN, S. A. *Correction and extension of the relationship of interpersonal perception to effectiveness in basketball teams*. Urbana, Ill.: Bureau of Research & Service, Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 5, Contract N6ori-07135.)

Received June 27, 1956.



# CASTRATION, CIRCUMCISION, AND ANTI-SEMITISM

IRVING E. ALEXANDER AND SHELDON BLACKMAN<sup>1</sup>

*Princeton University*

**I**N his writings, Freud gave recurrent expression to the view that fear of castration is a causal factor in anti-Semitism. In 1909 he wrote:

The castration complex is the deepest unconscious root of anti-Semitism; for even in the nursery little boys hear that a Jew has something cut off his penis—a piece of his penis, they think—and this gives them a right to despise Jews (4, p. 179).

In later statements he pointed out that castration and circumcision are equivalent in the unconscious (5), and also that the strangeness of the custom for gentiles serves to activate unconscious castration fear (6). That these views found some measure of acceptance in psychoanalytic circles can be seen in the work of Fenichel (2).

As a consequence of changing medical attitudes toward infant care, circumcision is no longer confined to one group in our culture. This turn of events provides an opportunity to gather some controlled information pertinent to Freud's speculation. The particular question we shall ask is whether circumcision is a relevant variable in anti-Semitism as measured in a contemporary male gentile population.

## METHOD

The measure of anti-Semitism used in this study is the 52-item A-S scale described by Levinson and Sanford (8). This scale was chosen because it has been standardized on a college population and some evidence is available for its reliability and validity. A further point in its favor for the present purpose is that it includes a "threatening" scale consisting of ten items related to Jews as powerful and possible overpowering figures. In the case of all the test items, a score of seven indicates strong agreement with the anti-Semitic position, a score of one, strong disagreement. If, as happened rarely, an item was omitted or for other reasons could not be scored, a value of four was arbitrarily assigned. Appended to the test was a data sheet on which relevant background information was recorded. Included was a section on health which contained questions on surgery undergone, e.g. appendectomy, tonsillectomy, circumcision, etc., and the time at which the operations had been performed.

This scale was administered to 381 students drawn from various psychology courses at Princeton and Rutgers Universities.<sup>2</sup> The subjects (Ss) included in this study are 142 gentile male undergraduates selected from the larger population and divided into two matched groups of 71 each. The groups were matched

<sup>1</sup> Now at the Ohio State University.

<sup>2</sup> We are grateful to Mr. Edward Nolan of Rutgers University, Newark Branch, for assistance in gathering the sample.

TABLE 1  
MEANS AND STANDARD DEVIATIONS OF SCORES ON THE A-S SCALE

Group	Mean	SD
Full Scale		
Circumcised	146.5	42.8
Uncircumcised	151.8	43.4
"Threatening" Scale		
Circumcised	26.4	7.5
Uncircumcised	27.4	9.4

on the basis of background variables the relationship of which to anti-Semitism had been investigated in the literature (7). These variables include family income, geographical location of home town, religion, church attendance, and frequency and intimacy of contact with Jews. The sample is relatively homogeneous with regard to other background variables such as age, rural-urban differences, and intelligence so that no effort other than inspection was made to control these factors. The one variable upon which the groups are known to differ systematically is whether or not the members are circumcised.

## RESULTS

Means and standard deviations for each group on both the full scale and the "threatening" subscale are presented in Table 1. In both instances, on the full scale and on the threatening scale, the group mean for the uncircumcised Ss exceeded that of the circumcised. However, these differences were small and statistically unreliable. An examination of the *F* ratios (full scale *F* = .43, threatening scale *F* = 1.01) gives no grounds for rejecting the view that they are attributable to chance.<sup>3</sup> The results of this study do not support the hypothesis that circumcision is a relevant variable in determining attitudes towards Jews and in turn cast doubt upon Freud's assumption as to the relation of castration anxiety to anti-Semitism.

## DISCUSSION

Several considerations come to mind in assessing the bearing that this study might have on the

<sup>3</sup> The analysis of variance technique does not take into account the positive correlation introduced by the matching of the groups. No satisfactory technique exists for estimating the correlation when groups are matched on the basis of total composition. The magnitude of *F* in each case is so small that the correlation would have to be above .9 in order to change the interpretation of our results.

castration-anti-Semitism hypothesis. One concerns the apparent evolution of attitudes toward this practice in the last twenty years. If few people in the culture at the present time, circumcised or not, have negative feelings about infantile circumcision, one might argue that this change has obscured the validity of Freud's original contention. The custom is no longer strange, and therefore ought not to arouse castration fear. This argument we would grant for children born in the present and future decades, but not for children born during the early stages of cultural change in attitude. During our Ss' formative years, both positive and negative attitudes toward circumcision were widespread and accepted in the larger culture.

Another consideration is whether we were dealing with a subculture so demanding of conformity from its members (the sample was predominantly composed of Princeton students) that it tended to obscure any real underlying attitudinal differences. That is to say, perhaps we were getting a reflection of a Princeton stereotype towards Jews rather than a true indication of the S's feelings. The fact that the scores of Jewish and gentile students in the entire Princeton sample differ significantly ( $p = .05$ ) suggests that the demands of the subculture leave room for expression of attitudinal differences.

Recently, two other investigators have presented discussions on the relationship of circumcision to castration which bear indirectly on the question we are considering. Bettelheim (1) and Fraenkel (3) have both questioned the widespread psychoanalytic belief that circumcision as a ritual stems from or has the meaning of a castrating act. Bettelheim, after a study of some anthropological and clinical material, finds no general evidence to support the idea that the practice arose in order to maintain an incest taboo or even that it evolved from father dominated societies. The Jewish practice, a relatively late one in history, he asserts is atypical in that it is carried out in infancy and is restricted entirely to males. The meaning of the circumcision rites, according to Bettelheim, is more frequently related to fertility and to the individual fulfillment of an unconscious bisexual wish. He offers no support for the assumption that circumcision and castration are necessarily symbolically equivalent or that circumcision would inevitably have negative valence in the unconscious.

Fraenkel concluded from a study of ancient

Jewish religious treatises, particularly the Pentateuch, that circumcision appeared to be more related to the benediction of fecundity than to an attenuated castration inflicted by the chief on his son. Only after circumcision was Abraham able to beget Isaac. Fraenkel characterizes the promise of the ritual by a quotation from the book of Genesis, "His seed will be as numerous as the sands of the earth." If the circumcision-castration argument cannot therefore be invoked with confidence, then Freud's speculation with regard to anti-Semitism loses its *raison d'être*.

#### SUMMARY

To obtain data relevant to Freud's speculations on the relationship between circumcision and anti-Semitism, attitudes towards Jews were measured among matched groups of circumcised and uncircumcised male gentile college students. No reliable differences in attitude were found between circumcised and uncircumcised subjects. These findings indicate the relative unimportance of circumcision as an isolated factor in anti-Semitism. The implications for the relationships involving circumcision, castration, and anti-Semitism are discussed.

#### REFERENCES

1. BETTELHEIM, B. *Symbolic wounds*. Glencoe, Illinois: Free Press, 1954.
2. FENICHEL, O. Elements of a psychoanalytic theory of anti-Semitism. In E. Simmel (Ed.), *Anti-Semitism*. New York: International Universities Press, 1946.
3. FRAENKEL, E. La circoncision chez les Juifs peut-elle s'expliquer comme une castration atténuée infligée à ses fils par le chef de la horde? *Psyché*, 1952, 7, 377-385.
4. FREUD, S. *Collected papers*. Vol. 3. An analysis of a phobia in a five year old boy. London: Hogarth Press, 1925.
5. FREUD, S. Eine Kindheitserinnerung des Leonardo Da Vinci. *Gesammelte werke*. Achter Band. London: Imago Publishing Company, 1943.
6. FREUD, S. *Moses and monotheism*. New York: Vintage Books, 1955.
7. HARDING, J., KUTNER, B., PROSHANSKY, H., & CHEIN, I. Prejudice and ethnic relations. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. II. Cambridge: Addison-Wesley, 1954. Pp. 1021-1061.
8. LEVINSON, D. J., & SANFORD, R. N. A scale for the measurement of anti-Semitism. *J. Psychol.*, 1944, 17, 339-370.

Received September 13, 1956.



# CONSISTENCY IN RESPONSE TO GROUP PRESSURES<sup>1</sup>

STANLEY ROSNER

*Society for Mental Hygiene, Bridgeport, Connecticut*

THE studies to be described take as their point of departure the investigations of Asch (1, 2) on modification of judgments under group pressure in face-to-face situations. Asch has reported that naive subjects (Ss) yielded approximately one-third of the time to the erroneous judgments of a tutored majority in matching the length of comparison lines with a standard line. The Ss who yielded early in the line series (consisting of 12 critical trials) tended to continue to yield throughout, while Ss who were independent of the announced judgments of the tutored majority, tended to continue independent throughout.

The purpose of the present studies is to answer the question: Are the consistencies in yielding or not yielding specific to the situation studied, i.e., judgment of lines, or do they extend beyond it? Is such consistency episodic or situationally determined, or is it a consequence of relatively settled modes of orientation?

## METHOD

To answer the above question, 88 student nurses<sup>2</sup> were observed in each of several Asch-type situations, in each of which they had to make a choice between being independent or conforming to the judgments of three accomplices<sup>3</sup> who were tutored by the experimenter to make erroneous judgments.

### Tasks<sup>4</sup>

Three memory tasks (nonsense words, nonsense figures, and a simple paragraph), a questionnaire, and the line series used by Asch (2) were employed in the experiment. Necessary controls for order of presentation and equating of control and experimental series were observed.

<sup>1</sup> This study is part of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy to the Graduate Faculty of Political and Social Science, The New School for Social Research, New York City. The writer owes a special debt of gratitude to Dr. Mary Henle for her encouragement of the research and for her critical reading of the manuscript.

<sup>2</sup> Thanks are owed to Miss Evelyn Lehrer, Assistant Director (formerly) of the Westchester School of Nursing and to Miss Barbara Tate, Associate Director of St. Lukes Hospital School of Nursing for helping to make their respective student bodies available.

<sup>3</sup> Asch (2) has shown that the effect of a majority of three on the judgments of naive Ss is as great as that of larger groups.

<sup>4</sup> For details, the reader is referred to the dissertation (3).

## RESULTS

1. Where comparisons are possible, present findings confirm those of Asch (1, 2).

2. A high degree of consistency in response to group pressure was found within a given task at a single sitting; Asch's finding that Ss who were independent early in the experimental session continued independent and Ss who yielded early in the experimental session continued to yield is substantiated with the present population.

3. Response to group pressure was significantly consistent between tasks at a single sitting. The Ss who were low yielders on one task tended to be low yielders on other tasks; high yielders on one task tended to be high yielders on other tasks. Phi coefficients ranging from .33 to .54 were obtained between pairs of tasks with respect to high and low yielders.

4. Response to group pressure was significantly consistent over a period of time. The Ss who yielded or did not yield during one sitting tended to yield or not to yield, respectively, during later sittings as shown in Table 1.

5. The embedded figures test (4, 5) was administered to 20 high and 20 low yielders to explore the relationship between this test and the group pressure situation. No significant difference was found between the two groups in the amount of time they required to locate the simple in the

TABLE 1  
CONSISTENCY OF YIELDING BEHAVIOR OVER A TIME LAPSE

Subjects	No. of Judgments Showing Yielding	
	Lines Session I*	Lines Session II
Yielders		
1	12	11
2	8	5
3	8	12
4	11	1
5	9	12
	48	41
Total		
Nonyielders		
1	0	1
2	0	0
3	2	1
4	2	0
5	0	1
	4	3
Total		

\* Procedure identical to that described by Asch (2).

complex figures, though the existing differences were in the expected direction. The yielders failed to locate the simple in the complex figures significantly more often than the nonyielders.

#### REFERENCES

1. ASCH, S. E. *Social psychology*. New York: Prentice Hall, 1952.
2. ASCH, S. E. Studies on independence and conformity [Pt. 1], a minority of one against a unanimous

majority. *Psychol. Monogr.*, 1956, 70, No. 9 (Whole no. 416).

3. ROSNER, S. Studies of group pressure. Unpublished doctor's dissertation, New School for Social Research, 1956.
4. WITKIN, H. A. Individual differences in ease of perception of embedded figures. *J. Pers.*, 1950, 19, 1-15.
5. WITKIN, H. A. *et al.* *Personality through perception*. New York: Harper, 1954.

Received September 28, 1956.

### ON THE RELATIONS BETWEEN DISCRIMINATION REACTION TIME, ANXIETY, AND INTELLIGENCE

G. ROBERT GRICE

*University of Illinois*

IN A test of a theory concerning the effects of drive on performance involving conflicting habits, Grice (4) found the predicted difference on a complex discrimination reaction time task between two groups differing in "manifest anxiety." However, a covariance analysis indicated that this difference could be attributed to differences between the groups in intellectual ability. Freeman and Maher (3) have criticized Grice's conclusion on the basis of speculations that the Discrimination Reaction Time Test (DRT) is "largely perceptual in character" and that perceptual factors in the Clerical Aptitude Index (employed as an intellectual measure) may have produced the obtained relation rather than general intellectual ability. Citing the excellent study by Fleishman and Hempel (2) as evidence, they argue that intellectual components in DRT tend to disappear with practice and that perceptual speed and simple reaction time assume greater importance.

In the first place, it should be pointed out that the task analyzed and described by Grice was quite different from that employed by Fleishman and Hempel. Instead of a long period of practice with a single set of instructions, the task (Grice's Problem 5) involved a signal, one second in advance of each reaction, which informed S which of four rather complicated sets of instructions were to be employed in determining the correct response. It would appear likely that this involved more intellectual ability than even the early trials of the conventional DRT task. In the second place, it should be pointed out that "Perceptual Speed" was the only purely perceptual factor which emerged in the Fleishman and Hempel analysis and that it accounted for only five per cent of the DRT variance even after long practice.

This cannot be regarded as strong evidence that the test was "largely perceptual in character."

The other point raised by Freeman and Maher concerned the suitability of the Clerical Aptitude Index as a general intellectual measure and the question of the presence of perceptual factors in it. They ask, rhetorically, "Were not Air Force data on tests more universally recognized as general intellectual measures available?" The answer is simple. No. The Clerical Aptitude Index is actually a rather good measure of intellectual ability as Grice originally stated. It is based on four tests: Word Knowledge, a test of academic vocabulary; Background for Current Affairs, a test of general information; Numerical Operations, a test involving simple computations; and Dial and Table Reading, a test requiring the verification of readings on groups of dials and the determination of certain information by reading mathematical tables. It was this last test which lead to the statement by Grice that the index contained perceptual factors. However, perceptual ability is by no means its major source of variance. In a factor analysis by Zachert (5), the test had a loading of .33 on a factor named Perceptual Speed. It also had loadings of .50 on Numerical Facility, .39 on General Reasoning, and .29 on Verbal Comprehension. In a more recent analysis by Bechtoldt (1), it had a loading of .17 on a Space factor and loadings of .31 and .26 on Academic Background and Induction respectively. The other tests show their major loadings on factors named Verbal Comprehension, Number, and Academic Background (1). All four tests have their major projections on second-order factors labeled by Bechtoldt as "G" (General Intellectual) and "VM" (Facility with Verbal Materials).

With the nature of the tests in mind, it can now



TABLE 1

PERFORMANCE OF HIGH AND LOW ANXIOUS GROUPS  
ON EACH TEST

Test	High A	Low A	<i>t</i>	<i>p</i>
	Mean	Mean		
Word Knowledge	4.40	6.03	3.47	<.01
Numerical Operations	4.40	5.37	2.05	<.05
Dial & Table Reading	3.37	4.70	2.51	<.02
Bkgd. Cur. Aff.	4.13	5.40	2.61	<.01

be pointed out that in Grice's study, each of the four significantly differentiated between the high and low anxious groups and that each was also positively correlated with response speed in the reaction time task. The means of the high and low anxious groups and the values of *t* are presented in Table 1. The product-moment correlations of response speed in Problem 5 in the DRT task with the individual tests were as follows: with Word Knowledge, .50; with Numerical Operations, .42; with Dial and Table Reading, .63; and with Background for Current Affairs, .33.

These considerations strongly suggest that the relationships reported by Grice were due to intel-

lectual variance in common with "manifest anxiety," "clerical aptitude," and the reaction time task, and that the speculations of Freeman and Maher are contrary to fact.

## REFERENCES

1. BECHTOLDT, H. P. Factor analysis of the Airman Classification Battery with civilian reference tests. USAF, Hum. Resour. Res. Cent., Res. Bull., 1953, No. 53-59.
2. FLEISHMAN, E. A., & HEMPEL, W. E. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, 49, 301-312.
3. FREEMAN, J. T., & MAHER, H. A note on "Discrimination reaction time as a function of anxiety and intelligence." *J. abnorm. soc. Psychol.*, 1957, 54, 134-135.
4. GRICE, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, 50, 71-74.
5. ZACHERT, VIRGINIA. Factor analyses of the Army, Navy, and Air Force classification batteries. USAF, Hum. Resour. Res. Cent., Res. Bull., 1952, No. 52-12.

Received March 7, 1957. Prior publication.

## SECURITY OF JUDGES AS A FACTOR IN IMPRESSIONS OF WARMTH IN OTHERS

JOSEPH BOSSOM AND A. H. MASLOW

Brandeis University

THE effect of variation among the people being judged has been considered in previous studies of first impressions (2). The present study is concerned with the effect of variation among judges upon their first impressions of others. Differences in the first impressions of secure and insecure judges are investigated.

## METHOD

## Subjects

Judges were 105 undergraduate students enrolled in an elementary course in psychology. From this total group were selected 22 judges who made scores of 9 points or less on the Maslow S-1 test for security (1), the secure group. Their scores ranged from 0.89 to 1.62 *SD* below the mean of the total group. Twenty-two judges, who scored 30 points or more comprised the insecure group. Their scores ranged from 0.71 to 2.65 *SD* above the total group mean.

## Stimulus

The "objects" to be judged consisted of the first 200 photographs obtained from yearbooks of former graduating classes. Except for difference in age (about two years), the "objects" and the judges come from about the same socioeconomic, racial, and religious background.

## Exposure to Stimulus

Each judge was handed an answer sheet with line space for 200 responses. The categories of response were headed: Very Warm; Warm; Cold; Very Cold. All judges were given the following verbal instructions:

You will be shown 200 photographs. They will appear on the screen in front of the room. These photographs will be changed rather rapidly. After looking at a photograph you will have sufficient time to indicate your impression of the person in the photograph. This is done by placing a mark in the appropriate box on your answer sheet.

Subsequently, the 200 photographs were projected on a screen. Each photograph was exposed for about 5 seconds. Periodically, the number of the photograph was called out so that judges could determine whether they had missed an item.

## RESULTS AND DISCUSSION

Judges in all groups were more likely to ascribe warmth than coldness to the persons represented in the photographs. The secure group, the insecure group, and the remaining judges (others) were compared with respect to the proportion in each group who ascribed warmth more frequently than coldness. As can be seen in Table 1, the secure

TABLE 1  
PERCENTAGE OF JUDGES IN EACH GROUP ASCRIBING  
WARMTH MORE OFTEN THAN  
COLDNESS TO PHOTOGRAPHS

Response Categories Compared	Group			Probability of Difference
	Secure (N = 22)	Insecure (N = 22)	Others (N = 61)	
Warm vs. Cold	95%	68%	78%	Sec. vs. Others .10 Insec. vs. Others .50 Sec. vs. Insec. .05
V. Warm & Warm vs. V. Cold & Cold	90%	54%	78%	Sec. vs. Others .80 Insec. vs. Others .05 Sec. vs. Insec. .02
V. Cold vs. V. Warm	77%	68%	42%	Sec. vs. Others .01 Insec. vs. Others .05 Sec. vs. Insec. .70

and insecure groups differed significantly ( $p = .05$ ) in the number of judges scoring Warm as compared with Cold. The difference between these groups was also significant when the Very Warm and Warm categories taken together were compared with Cold and Very Cold ( $p = .02$ ). In these comparisons a greater percentage of judges gave Warm responses in the secure group than among the "others." The "others," in turn, had a greater percentage of judges scoring warmth than the insecure group. When only the Very Warm and Very Cold categories were considered, significant differences were not obtained between

the secure and insecure groups. In addition, a higher percentage of judges in both extreme groups employ the Very Warm category than among the intermediate group of "others."

Since our data do not show whether the faces in the photographs are in "reality" warm or cold, we cannot attribute the obtained group differences to differential efficiency of perception of reality, projections based upon habitual expectations in interpersonal relationships, or to Freudian projection. Decision among these alternatives must await further research.

#### SUMMARY

From a larger group of judges, a secure and an insecure group were selected. All judges were exposed to a series of 200 photographs. Judges were required to indicate whether the person in the photograph was Very Warm, Warm, Cold, or Very Cold. The percentage of judges reporting Warm impressions more often than Cold impressions tended to be greater in the secure group than in the insecure group.

#### REFERENCES

1. MASLOW, A. H. A clinically derived test for measuring security-insecurity. *J. gen. Psychol.*, 1943, 33, 21-41.
2. TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, 52, 1-21.

Received August 27, 1956.



# PREDICTING THE DIRECTION OF SELECTIVE RECALL: ITS RELATION TO EGO STRENGTH AND ACHIEVEMENT<sup>1</sup>

THELMA G. ALPER

Wellesley College

WHEN the research findings in a given area continue over the course of many years to be as diverse as those on the recall of interrupted vs. completed tasks, the challenge to find out why can be strong. This has been the motivation behind the series of four papers by the writer (1, 2, 3, 4), the first of which appeared in 1946. At the time the research for the first two papers was undertaken, several studies had appeared. All of them had been designed to test further the conditions under which the Zeigarnik (29) finding, that interrupted tasks are more likely to be recalled than completed ones, would hold. Many of these studies, as previous reviews of the literature have shown (e.g., 1, 3, 13, 26), seemed not to support Zeigarnik since they did not yield statistically significant recall differences in the expected direction under conditions where completion and incompletion of tasks was experimentally controlled. One experiment, in fact, (24) indicated a reversal in the recall direction, i.e., a greater recall of completed than of interrupted tasks. The subjects (Ss) who served in that experiment, a group of crippled children, had been told "that they were to be given a test to determine how well they could do puzzles and that a prize would be given the one who did best" (24, p. 249). Not all of the Ss behaved in the same way. Some showed the reversal, some did not, others recalled an equal number of completed and interrupted tasks. The reversal pattern, however, was characteristic mainly of Ss rated by their teachers to be high on "pride." Though the ratings were admittedly not highly reliable, this was one of the first studies to support Zeigarnik's suggestion that personality factors may play a part in selective recall (29, p. 303). Zeigarnik's finding here with adults was that "ambitious" subjects, identified by remarks such as "What is the matter with me—!" or "I wouldn't have believed I could be so stupid," recalled

even *more* interrupted than completed tasks (29, p. 307).

The importance of another variable is also suggested by the Rosenzweig and Mason study; namely, the orientation of the experimental instructions. Working for a *prize* which you may get by doing well on a *test* is probably dynamically not the same as working on some tasks because someone asks you to do so.<sup>2</sup> Zeigarnik's instructions had been "I shall give you a series of tasks which you are to complete as rapidly and correctly as possible" (29, p. 300). In a later experiment, Rosenzweig (23) tested the orientation variable. He asked one group of college males, the "informal" group, to help the experimenter (E) classify some jigsaw puzzles in preparation for their use in a later experiment. Other Ss, the "formal" group, were told that the puzzles were part of an intelligence test. The selective recall results of the two groups were found to differ: the informal group recalled significantly more interrupted tasks, the formal group significantly more completed tasks. In Zeigarnik's study, no attempt had been made to maintain a uniform experimental atmosphere. On the contrary, Zeigarnik tells us that in keeping with differences in the Ss' reactions to the situation, wanting "to please the E," or "to excel as if in competition with others," or "interested in the task for its own sake" (29, p. 303), the E "did not preserve a fixed mien and method with all subjects. Those of the first type were allowed to see the experimenter's pleasure when the task was well done. Work by the second group was inspected with the air of an examiner, while the third group was allowed to work unmolested, the experimenter in this case remaining passive" (29, p. 303).

In the meantime, Alper (1), working also within Zeigarnik's tension theory of selective recall, and taking up Zeigarnik's (29, footnote 1, p. 312) suggestion that the recall of a given

<sup>1</sup> The writer acknowledges with thanks the invaluable help of Henry A. Murray, Robert W. White, and Eugenia Hanfmann.

<sup>2</sup> See Alper (3), for a further discussion of this point with special reference to the sample used in the Rosenzweig and Mason (24) study.

*S* might differ in different experimental settings, proposed that a more sensitive way to test for the conditions that influence the direction of recall might be to measure the recall of the *same Ss* under two different experimental settings: task orientation in one session, ego orientation in the other. To create a task-oriented atmosphere, the instructions emphasized, as Rosenzweig (23) had done, that *E* was testing the materials, not the *S*. For the ego-oriented atmosphere, where the intention was to threaten self-esteem, *Ss* were told that this was a standardized set of tasks and that they constituted a brief intelligence test devised for the army to use in selecting candidates for officers' training school. The *Ss* were all draft age college males. As predicted, *Ss* differed in their reactions to both instructions, and there were no significant recall differences under either orientation. When the recall results were correlated with personality ratings derived from an intensive clinical study of these *Ss* within the framework of Murray's (21) need theory of personality, however, two major patterns of recall emerged: (a) the Strong Ego pattern, consisting of the greater recall of interrupted tasks under task-oriented conditions, and of completed tasks under ego orientation; and (b) the Weak Ego pattern, consisting of a greater recall of completed tasks under task-oriented conditions, and of interrupted tasks under ego orientation.

One other variable had been controlled in the above study—the nature of the tasks used. Alper and Black (4) have shown that most of the tasks used by Zeigarnik (29), Lewis (15), and others, would be unacceptable to college students as IQ tests. Or, to put it concretely, failure to complete within a given time limit tasks such as bead-stringing, winding thread on a spool, circling vowels, etc., would not be regarded by an *S* as a reflection on his intelligence, though failure to complete an anagram or a scrambled phrases task would. The use of tasks not commensurate with the atmosphere *E* intends by his instructions may account for some of the previously equivocal findings in this field, as the writer has pointed out elsewhere (1, 4). In the experiments referred to above, the writer had used a scrambled phrases task.

To test the validity of the Strong and Weak Ego recall patterns, the writer designed another

experiment. Now two groups of *Ss*, a Strong Ego and a Weak Ego group, respectively, were selected in advance by means of specially constructed instruments. From the direction of their total recall patterns over two sessions, task-oriented first, and later an ego-oriented one, it was possible to predict correctly the group membership of these *Ss* as Strong or as Weak Egos significantly beyond chance. This experiment has been cited briefly elsewhere (2), but has not been reported in full.

In 1952, reacting to Sears' (26) recommendation that the interruption technique be abandoned as unsatisfactory since so little agreement had been reached by the various *Es*, the writer (3) tried to show: (a) that both orientation and personality were important in selective recall; and (b) that when both were considered, the seemingly diverse results of Zeigarnik (29), Rosenzweig and Mason (24), Rosenzweig (23), Glixman (13), Sanford (25), and others could be fitted into a common theoretical framework. The writer had intended to let the matter rest here. Recently, however, studies by Atkinson (5), and Atkinson and Raphelson (6) have renewed her interest since the data reported by these authors put a different interpretation on the results of earlier studies. They report that high *n* Achievers recall more interrupted than completed tasks under Achievement Orientation while low *n* Achievers show a reversed recall pattern. While these are the same recall patterns the writer has discussed, the personality relationships involved are not the same and the interpretations given them are also not the same. In fact, on the surface, at least, the *n* Achievement and Ego Strength frameworks would seem to have yielded dynamically opposite results.

Since *n* Achievement had not been measured in the writer's earlier studies, to test the relationship between *n* Achievement and Ego Strength requires a new experiment. This experiment has now been done and is reported in Part II of the present paper. In Part I, the writer has gone back to the prediction study mentioned above and is reporting it now in full since it confirms the validity of the Strong and Weak Ego recall patterns statistically revealed by the syndrome analysis used in the earlier study. It provides us, therefore, with a firmer base for testing the



relationship between *n* Achievement and Ego Strength and for applying the results to other selective recall studies.

## PART I

### PREDICTING THE DIRECTION OF RECALL OF STRONG AND WEAK EGO SUBJECTS

In the writer's earlier papers on the recall of interrupted vs. completed tasks (1, 2), the major premise had been that the direction of the recall ratio is a function of the personality structure of the individual recaller. It had been hypothesized "that the direction of selective recall of a given subject will differ in a non-self-esteem involving laboratory situation and in a laboratory situation where self-esteem is objectively threatened in a manner which is consistent with the self-esteem needs of that subject" (2, pp. 104-105). In designing the experiment to test this hypothesis, therefore, both instructional set and personality structure were taken into account.

Ten male undergraduates had served as Ss. Each was tested first in a task-oriented and later in an ego-oriented setting. The intention had been to threaten self-esteem objectively only in the latter setting. Since personality ratings, based on intensive clinical studies of these Ss within the framework of Murray's (21) need theory of personality were available, selective recall could be correlated with personality structure. Horn's (14) syndrome analysis technique was used for this purpose.

The analysis yielded two major selective recall patterns. One pattern, made up of the recall of more interrupted than completed tasks under Task Orientation and of more completed than interrupted tasks under Ego Orientation, was labeled the "Strong Ego" pattern since the syndrome analysis showed it to be characteristic only of Ss who had been rated high on *n* Recognition,<sup>3</sup> *n* Dominance,<sup>4</sup> Ego Strength,<sup>5</sup> and Conative Conjunctivity,<sup>6</sup>

<sup>3</sup> *n* Recognition: to seek to achieve something in order to win acclaim, renown, fame; to seek prestige and glory; to put oneself forward; to boast of one's accomplishments.

<sup>4</sup> *n* Dominance: to seek to control the sentiments and behavior of others by suggestion, seduction, persuasion or command; to lead others; to organize and direct groups of people; to get others to cooperate.

<sup>5</sup> Ego Strength: to know what one wants to do and has the capacity to do, and to do it. It is a matter of willing one's desires within the limits of one's abilities,

and low on Dejection Pessimism<sup>7</sup> and Ego-Ideal Intragression.<sup>8</sup> The opposite recall pattern, consisting of the recall of more completed than interrupted tasks under Ego Orientation, was labeled "Weak Ego" since it was characteristic only of Ss who had been rated high on Dejection-Pessimism and Ego-Ideal Intragression, and low on *n* Recognition, *n* Defendence,<sup>9</sup> *n* Counteractive Achievement,<sup>10</sup> and Narcism.<sup>11</sup>

In order to test whether these statistically derived patterns could be experimentally reproduced, the prediction study described here was designed.

### Method

Two groups of nine Ss each, a Strong Ego and a Weak Ego group respectively, were selected in advance of the experiment by means of specially constructed clinical tools. All Ss were tested under Task Orientation first and later under Ego Orientation.

It was predicted that the recall patterns of these Ss would be consistent with the findings of the earlier experiment since, apart from the fact that personality structure now served as an independent variable, the design of the prediction study followed closely the plan of the previous experiment.

*Selection of subjects.* Two procedures were used to select Ss. First, a self-rating scale, *The Psychological Insight Test*, was constructed to measure the personality variables in the Strong and Weak Ego syndromes. Second, Ss whose self-ratings identified them as Strong or as Weak Egos were interviewed. Only Ss whose

existing or acquirable. This manifests itself chiefly as a successful *n* Achievement, giving proof of the power to persist; an unconquerable will.

<sup>6</sup> Conative Conjunctivity: to organize one's efforts; to make plans and to follow them; to force drives into an efficient pattern; to live an ordered life. This variable gives evidence of the 'coordinating powers of the ego.'

<sup>7</sup> Dejection Pessimism: to worry about one's shortcomings; to be pessimistic about life; to recover slowly from disappointment.

<sup>8</sup> Ego-Ideal Intragression: to feel humiliated and ashamed after failure; to be burdened by feelings of inferiority. The ego is belittled and criticized for stupidity after failure.

<sup>9</sup> *n* Defendence: to defend oneself against blame, belittlement, assault; to justify one's actions; to offer extenuations, explanations and excuses for oneself.

<sup>10</sup> *n* Counteractive Achievement: to set difficult goals for oneself and attempt to reach them. To stick at a task until satisfied with the results; to overcome obstacles; to re strive after failure; to redouble one's efforts in the face of frustration.

<sup>11</sup> Narcism: to perceive the world from a personal or subjective view-point; to be disdainful of others; to be dominated by ruthless self-seeking; to purport to be indifferent to one's failures and to one's critics; to be self-satisfied; to be eccentric.

TABLE 1  
THE PSYCHOLOGICAL INSIGHT TEST

A Sample of the First Twelve Items

1. I enjoy acting before a large audience. (*n* Recognition)
2. I prefer to rely on my own judgments rather than depend on other men's opinions. (Narcism)
3. I look upon failure as a challenge to be met by redoubled efforts to succeed. (*n* Counteractive Achievement)
4. I am clearly conscious of my main goals in life. (Ego Strength)
5. I usually can think of something to say in my own defense. (*n* Defenceance)
6. I love good music and good wine and have become quite a connoisseur in such matters.
7. I usually decide for myself what I want to do and do it. (Conative Conjunctivity)
8. I am more apt to give in to an opponent than to keep on arguing with him. (Dejection Pessimism)
9. I spend a lot of time thinking about my shortcomings. (Ego-Ideal Intragression)
10. I enjoy organizing or directing the activities of a group—team, club, or committee. (*n* Dominance)
11. I like to amuse people by wisecracking or playing the fool. (*n* Recognition)
12. I feel that I am temperamentally different from most people. (Narcism)

self-ratings were confirmed by the clinical interviewer were asked to serve in the experiment.

*The Psychological Insight Test.* This test, consisting of two hundred statements, was constructed by the writer according to principles outlined by Murray and his co-workers (21). It provides a measure of the six personality variables within the Strong Ego syndrome and the six within the Weak Ego. Since three of the variables, *n* Recognition, Dejection-Pessimism, and Ego-Ideal Intragression, appear in both syndromes, only nine variables are included in the test. Each is measured by a set of 20 statements. Twenty additional statements which tap a variety of disparate needs were included but not scored. (See Table 12.)

The test, self-administered, requires *S* to rate himself on each statement on a +3 to -3 scale, the scale's zero point being the behavioral and emotional reactions of most men of *S*'s age, i.e., "the hypothetical average among college men." Thus, +3 represents *very much more often* than the average college man, -3 *very much less often*, +1 average but on the high side, -1 average but on the low side, etc.

<sup>12</sup> These statements were added both to disguise the nature of the test and to facilitate scoring. Nine variables and a "joker" variable repeated 20 times each allowed for the following scoring arrangement: statements measuring a given variable, for example *n* Recognition, were placed in the first, eleventh, twenty-first, etc., positions; a second variable, Narcism, in the second, twelfth, twenty-second, etc., positions. Since the statements for a given variable are sufficiently diversified in the test the scheme is not readily detected by *S*. This plan is demonstrated by the sample set of statements in Table 1. The need which each measured has been noted in parentheses following the statement.

When scoring the test, Strong Ego and Weak Ego scores are computed separately for each *S*. Each score consists of two parts: the algebraic sum of those variables which should be high for the given pattern, and the algebraic sum of those variables which should be low. A Strong Ego's score on Strong Ego variables, thus, would yield a +, - pattern, a Weak Ego's score on Weak Ego variables, a -, + pattern. Arithmetically, the scores on the Strong Ego variables can range from +240, -120 to -240, +120. The same is true for the Weak Ego variables.<sup>13</sup>

*The Psychological Insight Test* was administered to 140 male undergraduates enrolled in an introductory course in psychology at Harvard University under "take home" conditions. Eighty-six tests were returned. A Strong and a Weak Ego score was computed for each *S*, following the procedure described above.

In order to determine the range of Strong and of Weak Ego scores obtained by this sample, the bipart scores were combined into a single score by changing the sign of the second half of the score and adding the resulting sum to the first half of the score. For example, a score of +50, -25 would become +75, a score of -50, +25 would become -75. The scores of the eighty-six *Ss* on the Strong Ego variables ranged from +174 to -174; on the Weak Ego variables from +152 to -112. The *Ss* whose scores fell in the upper quartile of the Strong Ego distribution were selected as potential Strong Ego *Ss*. The *Ss* whose scores fell in the lower quartile of the Weak Ego distribution were selected as potential Weak Ego *Ss*. Of these, twelve Strong Ego scoring *Ss* and eleven Weak Ego scoring *Ss* agreed to come for an interview to the Harvard Psychological Clinic for the purpose of "possible participation in some experiments at the Clinic."

*The interview.* The interview, conducted either by Dr. Henry A. Murray or by Dr. Robert W. White, was structured around the Strong and Weak Ego syndrome variables, definitions and criteria for which were provided to the interviewer. Using a 0-5 point scale, the interviewer rated *S* on eight of the nine variables in these syndromes<sup>14</sup> and on the basis of these ratings labeled *S* as a "Strong Ego", a "Weak Ego" or "neither consistently Strong nor consistently Weak."

If the self-ratings and the interviewer's ratings were in agreement, either consistently "Strong" or consistently "Weak," *S* was asked by the interviewer to serve in the experiment.<sup>15</sup> In this way, the two groups

<sup>13</sup> To obtain the maximum Strong Ego score, *S* would have to rate himself +3 on all statements (80) pertaining to *n* Recognition, *n* Dominance, Ego Strength, and Conative Conjunctivity, and -3 on all statements (40) pertaining to Dejection-Pessimism and Ego-Ideal Intragression. It is to be noted that unlike the McClelland *et al.* (16) *n* Achievement variable, the present syndromes do not represent opposite ends of the same personality continuum. Only some of the variables appear in both syndromes.

<sup>14</sup> Through an oversight on the part of the writer, one variable, Narcism, had been omitted from the interview schedule.

<sup>15</sup> A partial measure of the sensitivity of *The Psychological Insight Test* is reflected by the fact that only five of the 23 *Ss* interviewed had to be rejected for the



of nine *Ss* each were selected. Their group membership "Strong" or "Weak" was not known to the writer until after the prediction scores had been computed at the end of the experiment.

**Materials.** The experimental materials were the same as those used in the earlier experiment (1). They consisted of two comparable sets of scrambled phrases to be arranged into meaningful sentences. Of the 12 sentences in each set, six were readily solvable within the two-minute time limit, three too difficult for control *Ss* to solve within the time limit, and three were unsolvable. Each solvable sentence permitted four alternative solutions.

**Procedure.** The procedure of the earlier experiment (1) was followed exactly. The atmosphere and instructions in Session I, Task Orientation, were informal and objectively non-self-esteem threatening since ostensibly *S* was merely helping *E* find out whether the materials would be suitable for a later experiment. The atmosphere and instructions in Session II, Ego Orientation, a week later, were formal and objectively threatening to self-esteem, sentence solution now being presented as "a brief intelligence test which the Clinic had devised for use by the Army in selecting officer training school candidates." To heighten the threat, moreover, a male accomplice, introduced as another *S*, worked in the same room on the same sentences and "successfully completed" all tasks within the time limits allowed.

At the end of the session, following a ten-minute interpolated projective test task,<sup>16</sup> incidental recall of the sentences was measured.

**Treatment of recall data.** In order to adjust for individual differences in number of sentences actually completed, two percentage recall scores were computed for each *S*: percentage of completed tasks recalled, and percentage of interrupted tasks recalled. This was done separately for each session. Since percentage scores are nonlinear, each was converted into angles by the arc sin√percentage transformation (27, Table 16.8, p. 449) before applying tests of significance.

Prediction of whether a given *S* fitted the Strong or the Weak category was based on his total recall score. The formula used for obtaining *S*'s total recall score was as follows:

Total Recall Score =  $(I_{n1} - C_1) - (I_{n11} - C_{11})$   
where  $I_{n1}$  refers to *S*'s transformed score for the recall of interrupted tasks in Session I,  $C_1$  to his transformed score for the recall of completed tasks in Session I,  $I_{n11}$  to his transformed score for the recall of interrupted tasks in Session II, and  $C_{11}$  to his transformed score for the recall of completed tasks in Session II.

To support the earlier results (2), the total recall score of a Strong Ego should be positive, a Weak Ego negative. The correctness of the prediction was tested by means of Fisher's (11) exact test.

To test whether the experimental groups differed in the direction of their recall within each orientation,

experiment. One *S* scored "Strong" on the self-rating scale and was rated "Weak" by the interviewer and four *Ss* did not consistently fit the interview criteria for either "Strong" or "Weak."

<sup>16</sup> The results of this test, *The Mind Reading Test*, are not included in this report.

TABLE 2  
PREDICTION OF STRONG AND WEAK EGO,  
SELECTIVE RECALL PATTERNS FOR TWO  
GROUPS OF MALE UNDERGRADUATES,  
A STRONG EGO GROUP AND A  
WEAK EGO GROUP

Category	Strong Ego Recall Pattern <sup>a</sup>	Weak Ego Recall Pattern <sup>b</sup>	<i>P</i> <sup>*</sup>
Expected	9	9	
Obtained	9	8	
Correctly Predicted	7	6	.045

<sup>a</sup> Based on total recall pattern over two sessions,  $(I_{n1} - C_1) - (I_{n11} - C_{11})$ ; to be counted as a Strong Ego pattern the score must be positive.

<sup>b</sup> Based on total recall pattern over two sessions,  $(I_{n1} - C_1) - (I_{n11} - C_{11})$ ; to be counted as a Weak Ego pattern the score must be negative.

<sup>\*</sup> Computed by means of Fisher's (11) exact test.

the Mann-Whitney *U* Test (16) was applied to the transformed percentage recall scores.

**Treatment of performance data.** Differences between the two groups of *Ss* in number of sentences solved, in number of alternative solutions per sentence achieved, in time to achieve a first solution, and in total number of sentences recalled were analyzed by means of analysis of variance.

## Results

**Prediction of group membership.** Nine of the 18 *Ss* had positive total recall scores, eight negative total recall scores, one a zero score. Since a zero recall score does not permit prediction of group membership, this *S* was omitted from the prediction analysis.<sup>17</sup> As shown in Table 2, the prediction of Strong Ego was found to be correct for seven of the nine positive total recall scores, the prediction of Weak Ego for six of the eight negative ones. Fisher's exact test yields a *P* value of .045 for these data. The null hypothesis that the direction of selective recall is unrelated to personality structure, as measured in this experiment, can therefore be rejected.

**Selective recall of Strong and Weak Ego Groups within each orientation.** The analyses from this point on are based on the results obtained by the *Ss* within the two experimental groups, i.e., the nine Strong and the nine Weak Ego *Ss* as selected by the combined criteria of the self-rating scale and interview, described above.

Differences in percentage recall scores ob-

<sup>17</sup> This score was subsequently identified as the score of a Strong Ego *S*.

TABLE 3  
MEAN RECALL SCORES OF STRONG EGO AND  
WEAK EGO MALE SUBJECTS TESTED UNDER  
BOTH TASK-ORIENTED AND  
EGO-ORIENTED CONDITIONS\*

	Task-Orientation		Ego-Orientation	
	In	C	In	C
Strong Ego Ss	.37	.57	.26	.46
Weak Ego Ss	.29	.48	.31	.32
	<i>P</i> *		<i>P</i> *	
In	.07		.25	
C	.27		.06	
In—C	.45		.01	

\* Percentage recall scores converted into angles by the arc sin√percentage transformation (27, Table 16.8, p. 449).

\* Computed by means of Mann-Whitney *U* Test (14).

tained by the Ss in the two experimental groups are shown in Table 3. Under Task Orientation, the Strong Ego group tends to recall more interrupted tasks than the Weak Ego Ss ( $P = .07$ ). The Weak Ego Ss recall more completed tasks than do the Strong Ego Ss, but the difference is not statistically significant ( $P = .27$ ). There is also no difference in the so-called Zeigarnik effect, In — C ( $P = .45$ ), though the difference is in the expected direction.

Under Ego Orientation, Weak Ego Ss recall more interrupted tasks than do Strong Egos, but the difference is not statistically significant. Strong Ego Ss tend to recall more completed tasks than do Weak Egos, and this difference approaches significance ( $P = .06$ ). The difference in recall as measured by In — C for the two groups is significant ( $P = .01$ ). In other words, the discrepancy in favor of the greater recall of completed than of interrupted tasks is greater for the Strong Ego Ss than for the Weak Ego Ss.

*Performance of the Strong and Weak Ego Groups within each orientation.* Tables 4, 5, and 6 summarize the group differences for the performance variables. Table 4 shows that under Task Orientation, the two groups are equally efficient at solving the sentences. Under Ego Orientation, both groups lose in efficiency, the Strong Ego Ss losing less than the Weak Ego Ss, but the differences are not significant.

Since each solvable sentence permitted four solutions, a more sensitive measure of performance loss from Task Orientation to Ego

Orientation is afforded by the mean number of alternative solutions achieved under each condition. As Table 5 shows, the mean for Weak Ego Ss under Task Orientation is higher than for Strong Egos, but the difference is not significant. Under Ego Orientation, both groups show a drop. When differences between the two sessions are analyzed, the loss for the Weak Ego Ss is found to be significantly greater than for the Strong Ego Ss ( $P = .05$ ).

The poorer performance of Weak Ego Ss, as the instructions become more ego threatening, may result from increased rigidity, i.e., having once found a solution they are less able than Strong Egos to disturb that structure and to attain an alternative solution. The possibility of greater rigidity in Weak Egos is supported by the data concerning time to achieve a first solution. These data, summarized in Table 6, show no significant difference between groups in time required to achieve a first solution. In other words, Weak Ego Ss are not slower to find an initial solution to the problem when the instructions become more ego threatening (Table 6), but are significantly less able to restructure and achieve additional solutions within the two-minute time limit per sentence (Table 5).

As Table 7 shows, there is no significant overall memory loss as the conditions become more threatening to self-esteem. The differences within each experimental session, moreover, do not even approach significance.

The selective recall differences reported above, therefore, represent qualitative differences as a function of ego strength, not quantitative differences in total memory. The performance data indicate both qualitative and quantitative group differences. Under objectively non-self-esteem threatening conditions, the performance of the two groups is the same. When self-esteem is objectively threatened, however, Weak Ego Ss show the greater loss in efficiency.

## PART II

### THE RELATIONSHIP BETWEEN EGO STRENGTH AND *N* ACHIEVEMENT

In order to fit the data of Part I into the results of recall studies done within the McClelland *et al.* (18) framework of achievement motivation, the relationships between *n* Achievement as a single parameter of per-



TABLE 4

ANALYSIS OF VARIANCE FOR NUMBER OF SENTENCES SOLVED BY STRONG AND WEAK EGO COLLEGE MALES UNDER TWO EXPERIMENTAL CONDITIONS

UNDER TWO EXPERIMENTAL CONDITIONS														
Task Orientation					Ego Orientation					Difference Between Orientations (Ego - Task)				
	<i>df</i>	Sum of Squares	Var- iance	<i>F</i>		<i>df</i>	Sum of Squares	Var- iance	<i>F</i>		<i>df</i>	Sum of Squares	Var- iance	<i>F</i>
Total Between Ss Within	17	10.3			Total Between Ss Within	17	24.3			Total Between Ss Within	17	27.1		
	1	.1	.1	.16		1	4.5	4.5	3.66		1	5.5	5.5	4.07
	16	10.2	.64			16	19.8	1.23			16	21.6	1.35	
	Strong Egos		Weak Egos			Strong Egos		Weak Egos			Strong Egos		Weak Egos	
Mean	5.3		5.4		Mean	5.1		4.1		Mean	-0.2		-1.3	

TABLE 5

ANALYSIS OF VARIANCE FOR NUMBER OF ALTERNATIVE SOLUTIONS BY STRONG AND WEAK EGO COLLEGE MALES UNDER TWO EXPERIMENTAL CONDITIONS

UNDER TWO EXPERIMENTAL CONDITIONS

Task Orientation					Ego Orientation					Difference Between Orientations (Ego - Task)				
	df	Sum of Squares	Var- iance	F		df	Sum of Squares	Var- iance	F		df	Sum of Squares	Var- iance	F
Total Between Ss Within	17 1 16	133.8 8.1	8.1	1.03	Total Between Ss Within	17 1 16	73.8 5.6 68.2	5.6 4.3	1.3	Total Between Ss Within	17 1 16	187.5 44.5 143.0	44.5 8.9	5.0*
	Strong Egos		Weak Egos			Strong Egos		Weak Egos			Strong Egos		Weak Egos	
Mean	8.5		9.8		Mean	6.7		5.6		Mean	-1.8		-4.2	

\* Indicates significance beyond the .05 level of confidence.

TABLE 6

ANALYSIS OF VARIANCE FOR AVERAGE TIME IN SECONDS TAKEN BY STRONG AND WEAK COLLEGE MALES TO ATTAIN A FIRST SOLUTION UNDER TWO EXPERIMENTAL CONDITIONS

TO ATTAIN A FIRST SOLUTION UNDER TWO EXPERIMENTAL CONDITIONS														
Task Orientation					Ego Orientation					Difference Between Orientations (Ego - Task)				
	df	Sum of Squares	Variance	F		df	Sum of Squares	Variance	F		df	Sum of Squares	Variance	F
Total Between Ss Within	17 1 16	3766.54 21.35 3745.19	21.35 234.07	.01	Total Between Ss Within	17 1 16	7212.72 90.60 7122.12	90.60 455.13	.20	Total Between Ss Within	17 1 16	5304.37 23.35 5281.02	23.35 330.06	.07
	Strong Egos		Weak Egos			Strong Egos		Weak Egos			Strong Egos		Weak Egos	
Mean	68.7		70.9		Mean	78.3		82.7		Mean	+9.6		+11.9	

sonality and the personality syndromes Ego Strength and Ego Weakness have to be examined. Since *n* Achievement was not measured in the prediction study, a new experiment

was designed. The specific question it asks is: Do Ss who score high on *n* Achievement, as measured by the TAT method of McClelland *et al.* (18), more often fit only one of Alper's

TABLE 7  
ANALYSIS OF VARIANCE FOR NUMBER OF SENTENCES RECALLED BY STRONG AND WEAK EGO COLLEGE MALES  
UNDER TWO EXPERIMENTAL CONDITIONS

Task Orientation					Ego Orientation					Difference Between Orientations (Ego - Task)				
	df	Sum of Squares	Var- iance	F		df	Sum of Squares	Var- iance	F		df	Sum of Squares	Var- iance	F
Total Between Ss	17	30.9			Total Between Ss	17	54.9			Total Between Ss	17	61.8		
Within	1	2.7	2.7	1.5	Within	1	1.3	1.3	.38	Within	1	.2	.2	.05
	16	28.2	1.8			16	53.6	3.4			16	61.6	3.85	
		Strong Egos	Weak Egos				Strong Egos	Weak Egos				Strong Egos	Weak Egos	
Mean		5.4	4.7		Mean		4.2	3.7		Mean		-1.2	-1.0	

two personality syndromes, and Ss who score low on *n* Achievement, the other?

Although the Ss in this experiment were college females, rather than males, the sex factor should not affect the relationships being studied here since: (a) the design of the present experiment takes into account the conditions that should assure a high scoring *n* Achievement group; and (b) there are no reports of sex differences in the recall of interrupted vs. completed tasks. Indeed, in the Atkinson and Raphelson (6) study, the authors included the recall results of 22 female Ss "to add stability to the recall scores of men" (6, p. 352). They do not report separate recall results for the female sample.

Sex differences have been reported, however, in studies of *n* Achievement. McClelland *et al.* summarize them as follows: "(1) women get higher *n* Achievement scores than men under neutral conditions—two studies; (2) Achievement-Oriented does not increase *n* Achievement in women—three studies; (3) women's *n* Achievement scores seem as valid as men's in that they relate to performance in the same way; (4) the failure to find an increase in *n* Achievement score from achievement arousal is probably not due to the fact that the scores are already at a maximum" (18, p. 178). The clue for arousing higher *n* Achievement scores in women is seen by McClelland *et al.* to come from a study by Field (10) who reported that *n* Achievement in women is tied up with social acceptability rather than with reference to leadership capacity and intelligence. In Field's study, for example, females who were rated by their classmates as "liked" scored

higher on *n* Achievement than those rated as "disliked," the ratings having been made after *n* Achievement had been measured and without advance knowledge that such ratings would be asked for. McClelland *et al.* conclude from these data that "if you want to arouse *n* Achievement in women, refer, as Field did, to their social acceptability; if you want to arouse *n* Achievement in men, refer, as we did to their leadership capacity and intelligence" (18, p. 181). Since Morrison (20) has subsequently confirmed Field's thesis by finding that female college leaders score higher on *n* Achievement than do female nonleaders, we may paraphrase McClelland *et al.*, as follows: if you want to be assured of a high *n* Achievement scoring female sample, use college leaders as Ss. Such students formed the experimental group in the present study.

An added advantage of a leader group for present purposes is the relationship reported by Atkinson and Raphelson (6) between selective recall and *n* Affiliation ("motivation for social acceptance" (6, p. 350)). Under a very relaxed orientation, Ss who scored high on *n* Affiliation, as measured by the TAT technique, recalled more interrupted than completed tasks, while Ss who scored low on *n* Affiliation showed the opposite recall pattern. Both Atkinson (5) and Atkinson and Raphelson (6) report that Ss who score high on *n* Achievement under Achievement Orientation, recall more interrupted tasks than do Ss who score low on *n* Achievement. These findings taken together suggest the possibility that Ss who score high on *n* Affiliation and low on *n* Achievement might show the recall pattern of Alper's



Strong Ego Ss; namely, greater recall of interrupted tasks under Task Orientation, of completed tasks under conditions which objectively stress that *S*'s performance is being measured. Similarly, Ss who score low on *n* Affiliation and high on *n* Achievement might be expected to show the reverse recall pattern characteristic of Alper's Weak Ego Ss. The fact that the recall of completed tasks has been found in the past to be a function of increased experimental stress, not of the strength of *S*'s *n* Achievement, may function against the hypothesized relationships between Ego Strength, *n* Affiliation, *n* Achievement, and recall pattern. On the other hand, since Atkinson and Raphelson (6) do not report how high *n* Affiliation Ss scored on *n* Achievement, though both measures were obtained for each *S*, nor have they tested the selective recall of the same Ss under both Task and Achievement Orientation, there is at present no clear contrary evidence against the hypothesis.

### Method

*The Ego Strength measure.* The Psychological Insight Test, as described in Part I, was used to measure Ego Strength. Designed originally for use with college males, it was revised for female Ss only by changing "male" to "female," "men" to "women," "his" to "her," etc., in the directions and in the body of the scale. As in Part I, it was taken under "take home" conditions. In the instructions, *E* attempted to keep the conditions as nonthreatening to self-esteem as possible by emphasizing both that names would be removed after coding and that the results would be used as part of a study of college students in general, not as tests of the participating individuals.<sup>18</sup>

The algebraic scoring scheme, as described in Part I, was used to determine *S*'s score on each of the nine variables measured. To fit the Strong Ego pattern, *S*'s self-ratings would involve positive scores on *n* Recognition, *n* Dominance, Ego-Strength and Conative Con-junctivity, negative scores on Dejection-Pessimism and Ego-Ideal Intragression. To fit the Weak Ego pattern, *S*'s self-ratings would involve negative scores on *n* Recognition, *n* Defendance, *n* Counteractive Achievement and Narcism, positive scores on Dejection-Pessimism, and Ego-Ideal Intragression. The Strong Ego, thus, would have a +, - pattern of scores, the Weak Ego a -, + pattern. A +, + pattern can be regarded as an inconsistent Strong Ego pattern, a -, - pattern an inconsistent Weak Ego pattern.<sup>19</sup>

<sup>18</sup> It is quite likely that such instructions would be more ego-threatening to some, less to others. To label them nonthreatening to self-esteem, then, is to stress *E*'s intent, not *S*'s involvement.

<sup>19</sup> Quantitative scores were obtained, as well as this pattern score, by following the scheme described in

*The n Achievement measure.* The choice of picture stimuli for measuring *n* Achievement was influenced by: (a) the finding by Veroff *et al.* (28) that pictures of male characters yield higher *n* Achievement scores than pictures of female characters for both male and female college Ss; (b) the writer's clinical experience with the usefulness of particular TAT pictures for evoking achievement needs; and (c) an interest in measuring parent-child relationships as perceived by high and low scoring Ss on the *n* Achievement variable. These criteria resulted in the selection of four pictures: Cards 1 and 7GF from *The Thematic Apperception Test* (22), i.e., boy and violin, and young girl and mother-figure, respectively, and Cards 10G and 11G from *The Michigan Picture Test* (19), i.e., young girl and father-figure, and young girl in empty classroom. The *n* Achievement scores reported here are based only on two of these, the boy and violin, which are listed among the McClelland *et al.* (18) stimuli for measuring *n* Achievement, and the girl in classroom which can be considered a female counterpart of their Card H, "a boy in checked shirt at desk, an open book in front of him" (18, p. 375). Since the young girl is typically seen as about ten years old, as is also the case for the boy in the the boy and violin picture, when the storyteller mentions age at all, these two pictures are particularly well matched for tapping reactions to sex differences of the stimuli.

To facilitate group presentation of the pictures, permission was obtained from the respective publishers to make the necessary slides as well as certain changes in the stimuli. For example, in Card 11G, the background of the empty classroom was blurred, giving the background less "an empty classroom" appearance.

*Experimental design.* Three groups of college females were used as Ss in this experiment, an experimental group and two control groups. The experimental group was comprised of 28 upperclassmen, all of whom had been elected to high office by their peers. Achievement motivation was tested by a college senior, herself a college leader.<sup>20</sup> The conditions of testing and the instructions were casual, the intention being merely to task-orient these Ss. To this end, the group was told by the student *E* that "some studies of college students were being made and that they had been asked to help." A second group, 24 upperclassmen in an advanced course in psychology, served as a sample of upperclass nonleaders. Office holders in the group were eliminated from the sample. The test was administered by the instructor in the course immediately following the completion of an hour examination in the course in which these students were enrolled. The instructions emphasized that the test measured personality and that it would be scored by the instructor. This group, therefore, was an ego-oriented group and was tested immediately following a real-life achievement situation under ego-oriented conditions. The third group was

Part I. Only pattern scores are considered here since the concern is merely with presence or absence of a given self-rating pattern in relation to the *n* Achievement variable.

<sup>20</sup> The writer wishes to thank Miss Maud Hazeltine, President of Student Government at Wellesley College in 1955-56, for administering the test to the leader group.

comprised of 24 students enrolled in an introductory course in psychology. These Ss, largely sophomores, constituted another nonleader sample. Their achievement motivation was measured during a regular class hour under casual task-oriented conditions. The instructions emphasized that this was merely "a demonstration of a psychological technique." Since demonstrations were frequent in the course, this device was used for task-orienting these Ss.

In all three groups, the test was administered under group testing conditions. The McClelland *et al.* (18, p. 98) instructions and questions followed the special instructions, task- or ego-oriented, for the given group, as noted above.

The following predictions were made:

1. Following Morrison (20), the mean *n* Achievement score of the task-oriented leader group will be significantly higher than the mean *n* Achievement score of the ego-oriented nonleader group and of the task-oriented nonleader group;

2. Following Veroff *et al.* (28), the mean *n* Achievement score of the ego-oriented nonleader group will not differ significantly from the mean *n* Achievement score of the task-oriented nonleader group.

3. Following Atkinson (5) and Atkinson and Raphelson (6), in the leader group only, Ss whose self-ratings fit the Strong Ego pattern will more often score low on *n* Achievement than high on *n* Achievement, while Ss whose self-ratings fit the Weak Ego pattern will more often score high on *n* Achievement. The prediction was limited to the leader group since only this group meets the criterion of high *n* Affiliation required by the Atkinson and Raphelson (6) recall findings as well as Morrison's (20) *n* Achievement findings for female Ss.

## Results

Scoring of Ego Strength and of *n* Achievement was done by the writer. Since the Ego Strength measure entails merely arithmetical addition of Ss' self-ratings, it presents no problem with respect to scorer reliability, a problem which arises, however, with the *n* Achievement measure. Two steps were taken to check on the reliability of the writer's scoring of *n* Achievement. First, several months after the initial self-instruction period with the McClelland *et al.* scoring categories (18, Chap. IV), the writer rescored the first story of each of the 30 Ss included in their Appendix I. A rank-order correlation test of the writer's scores checked against the scores reported for these stories by McClelland *et al.* (18) in their Appendix II yielded a rho value of  $+ .85$ . Secondly, to check the writer's self-consistency for scoring *n* Achievement, she rescored the stories told by the leader group one month after her original scoring. The rho value here was  $+ .96$ . Since these

TABLE 8  
MEAN *n* ACHIEVEMENT SCORES OF COLLEGE  
FEMALES UNDER VARIOUS ORIENTATIONS  
(Stimuli Were TAT Card I and Michigan  
Picture Test Card 11G)

	College leaders, task-oriented	Nonleaders, ego-oriented	Nonleaders, task-oriented
Mean <i>n</i> Achievement Score	3.1	4.1	3.5

	<i>t</i>	<i>df</i>	<i>P</i>
College leaders — non- leaders, ego-oriented	-.85	50	.50 — .40
College leaders — non- leaders, task-oriented	-.34	50	.80 — .70
Nonleaders, ego-oriented — nonleaders, task-oriented	-.55	46	.60 — .50

values of rho compare favorably with other reliability tests cited by McClelland *et al.* (18, pp. 185-187), it is assumed that the scoring is reliable.

*Analysis of the total mean n Achievement scores based on stories told to two pictures* (TAT Card 1 and Michigan Picture Test Card 11G). The mean *n* Achievement scores for the three groups of Ss are shown in Table 8. The data indicate that the three groups do not differ on *n* Achievement. The first prediction, that the leader group would score higher than the nonleader groups, is therefore not supported. In fact, on an absolute basis, the leader group scored lowest, the ego-oriented nonleader group scored highest, the task-oriented nonleader group in the middle, though none of the differences are significant.

Since the mean *n* Achievement scores of the two nonleader groups do not differ, the second prediction is supported. The fact that one group was ego-oriented during the testing and the other task-oriented, in other words, does not affect their *n* Achievement scores.

*Analysis of the effect of the sex of the character in the picture stimulus on n Achievement scores.* To check the possibility that the three groups may have been responding differently to the two pictures, separate *n* Achievement scores were computed for each picture. Dependent *t* tests, comparing the scores on the two pictures within a group, yielded the results summarized in Table 9. These data indicate that the two pictures are not equally effective in arousing *n* Achievement for all groups. The male picture arouses significantly more *n* Achievement



ment in the leader group than does the female picture, the mean of the former being 2.6, of the latter 0.7 ( $P < .05$ ). The differences are in the same direction for the ego-oriented nonleader group, but they are not significant ( $P < .20$ ). For the task-oriented nonleader group there is a reversal: the female picture tends to yield a higher mean *n* Achievement score than the male picture ( $P < .10$ ).

In view of these findings, the first two predictions were tested for each picture separately. The results, analyzed by means of independent *t* tests, are summarized in Table 10.

As is shown in the left half of Table 10, Prediction 1 is partly supported for the male picture: *n* Achievement is significantly higher for the leader group than for the task-oriented nonleader group ( $t = 7.96$ ;  $P < .001$ ), but there are no differences between the leader group and the ego-oriented nonleader group ( $t = -.53$ ;  $P < .60$ ).

Prediction 2 is *not* supported since the mean *n* Achievement score of the ego-oriented nonleader group is significantly higher than the mean *n* Achievement score of the task-oriented nonleader group ( $t = 2.04$ ;  $P = .05$ ).

The right half of Table 10 summarizes the

results for the female picture. Again the leader group does not differ from the ego-oriented nonleader group ( $t = -.56$ ;  $P < .60$ ) but does differ from the task-oriented nonleader group. The female picture yields a significantly lower mean *n* Achievement score for the task-oriented leader group than for the task-oriented nonleader group ( $t = -2.16$ ;  $P = .05$ ). The two nonleader groups also differ on the female picture, the ego-oriented nonleader group scoring significantly lower on *n* Achievement than the task-oriented nonleader group ( $t = -5.43$ ;  $P < .001$ ).

In brief, these results indicate that: (a) the sex of the character in the picture stimulus does affect the strength of the *n* Achievement score of college females; (b) *n* Achievement is more likely to be aroused in a female *S* by a male than by a female character if *S* either values social acceptance or is being tested in an achievement atmosphere; (c) *n* Achievement is more likely to be aroused in a female *S* by a female than by a male character if *S* does not value social acceptance and she is being tested in a task-oriented atmosphere.

*Relation between Ego-Strength patterns and n Achievement.* The Psychological Insight Test scores were labeled Strong Ego, Weak Ego, Not Consistently Strong, or Not Consistently Weak according to whether the score fitted a +, -; a -, +; a +, +; or a -, - pattern respectively, as described earlier. The frequency of these patterns within the leader and nonleader groups is shown in Table 11. The two nonleader groups were combined here since a chi-square analysis revealed them not to be different ( $\chi^2 = .75$ ;  $P < .90$ ).

In all groups, the Weak Ego pattern occurs rarely. The Strong Ego pattern occurs most frequently, but not more frequently than would be expected by chance alone when

TABLE 9

MEAN *n* ACHIEVEMENT SCORES OF COLLEGE FEMALES BASED ON STORIES TOLD TO A MALE AND TO A FEMALE PICTURE STIMULUS

Group	Male Picture	Female Picture	<i>t</i>	<i>df</i>	<i>P</i>
College leaders, task-oriented	2.6	.7	2.42	27	.05-.02
Nonleaders, ego-oriented	3.0	1.1	1.47	23	.20-.10
Nonleaders, task-oriented	1.3	2.3	-1.92	23	.10-.05

TABLE 10

MEAN ACHIEVEMENT SCORES OBTAINED BY COLLEGE FEMALES ON STORIES TOLD TO A MALE AND TO A FEMALE PICTURE STIMULUS

Group	Male Picture			Female Picture		
	<i>t</i>	<i>df</i>	<i>P</i>	<i>t</i>	<i>df</i>	<i>P</i>
College leaders, task-oriented—nonleaders, ego-oriented	-.53	50	.60-.50	-.56	50	.60-.50
College leaders, task-oriented—nonleaders, task-oriented	2.04	50	.05	-2.16	50	.05
Nonleaders, ego-oriented—nonleaders, task-oriented	7.96	46	<.001	-5.43	46	<.001

TABLE 11  
PATTERNS OF PSYCHOLOGICAL INSIGHT TEST  
SCORES OBTAINED BY FEMALE COLLEGE  
LEADERS AND FEMALE COLLEGE  
NONLEADERS

Group	N	Psychological Insight Test Patterns			
		Strong Ego	Weak Ego	Not Consistently Strong	Not Consistently Weak
Leaders	28	16	2	4	6
Nonleaders	46	21	6	15	4

TABLE 12  
RELATIONSHIP BETWEEN EGO STRENGTH AND  
n ACHIEVEMENT OF FEMALE COLLEGE  
STUDENTS  
(Analysis by chi square<sup>a</sup>)

	N	No. Scoring Above the n Achievement Mean of the Total Group	No. Scoring Below the n Achievement Mean of the Total Group	$\chi^2$	P <sup>a</sup>
Strong Ego Ss	16	3	13	6.31	<.02
Weak Ego Ss	2	0	2	—	—

<sup>a</sup> The correction for continuity was applied to these data.  
<sup>a</sup>  $df = 1$ .

Strong Ego frequency is compared with the frequency with which all other patterns combined occur.

Table 12 shows the number of Strong Ego Ss in the leader group whose *n* Achievement scores fell above the group's mean, and the number whose scores fell below the group's mean. Similar data are included for the Weak Ego Ss.

The Strong Ego data indicate that significantly more Ss in the leader group whose self-ratings fit the Strong Ego pattern score below the mean of the group on *n* Achievement ( $P = .02$ ). Thus, the prediction that Strong Ego Ss in the leader group would more often score low on *n* Achievement, not high, is supported.

The prediction that Weak Ego Ss in the leader group would score high on *n* Achievement was not tested since the pattern appeared so infrequently in this group.

A prediction of how Strong Ego Ss in non-leader groups would score with respect to their group means was not made since the *n* Achievement literature provided no basis for

making a prediction. The fact that in both nonleader groups as many Strong Ego Ss scored above their group mean as below it, however, strengthens the importance of the the inverse relationship between *n* Achievement and Ego Strength found in the leader group data.

## DISCUSSION

The results reported in Part I of this paper support the position taken earlier by the writer (1, 2, 3) that selective recall is lawfully related to the personality structure of the recaller. In the present study, two personality patterns, the Strong Ego and the Weak Ego, as previously (2) described, were studied further by selecting in advance of the experiment Ss who fitted these patterns. All Ss were tested under two experimental conditions, Task Orientation first, followed by Ego Orientation a week later. It was found that correct prediction of group membership on the basis of the selective recall pattern alone was possible significantly beyond chance: the Strong Ego pattern consisting of greater recall of completed than of interrupted tasks as the experimental conditions become more threatening to self-esteem; the Weak Ego pattern, of greater recall of interrupted than of completed tasks.

As in previous papers (2, 3), these results are interpreted to mean that Strong and Weak Ego Ss protect self-esteem in different ways. In situations in which completion is likely to be experienced as personal success, interruption as personal failure, i.e., when tasks are being performed under ego-oriented conditions, the Strong Ego is not overwhelmed by the failure, the Weak Ego is. The Strong Ego supports his self-esteem by focusing on the tasks he *did* complete, i.e., recalls more completed than interrupted tasks. This mechanism of defense is not available to the Weak Ego. Instead, the Weak Ego utilizes an ego-offensive mechanism (2): he focuses on his failures, i.e., recalls more interrupted than completed tasks, admitting his failures before others face him with having failed. Both types of recall, then, serve to reduce ego tensions under conditions experimentally designed to arouse ego, rather than task, tensions.

Similar recall patterns, but interpreted in very different ways, have been reported recently by Atkinson (5) and Atkinson and



Raphelson (6). The data of Part II support a reinterpretation of their findings. As a result, the *n* Achievement and Ego Strength studies can be shown to have yielded uniform patterns of selective recall.

Atkinson and Raphelson (6) have reported that *Ss* who score high on *n* Affiliation (motivation for social acceptance) as measured by the TAT technique of McClelland *et al.* (18) under the "very relaxed" orientation, recall more interrupted tasks than do *Ss* who score low on *n* Affiliation. As the conditions become more achievement oriented, the direction of selective recall correlates with *n* Achievement, not with *n* Affiliation. The *Ss* who score high on *n* Achievement tend to recall more interrupted than completed tasks, while *Ss* who score low on *n* Achievement recall more completed than interrupted tasks (5, 6). They interpret these results to mean that the *S* who recalls more interrupted tasks under relaxed conditions is reacting merely with task tensions. He wants to *please* the *E* and hence, like Zeigarnik's (29) *Ss*, when a task tension cannot be resolved, because *E* has interrupted *S* on the task, the tension persists, making interrupted tasks more available for recall than completed ones. The greater recall of interrupted than of completed tasks by high *n* Achievers, and the reverse pattern by low *n* Achievers, under achievement-oriented conditions are explained as follows: high *n* Achievers have a "positive disposition to excel—are predominantly success-oriented and have as their goal the feeling of satisfaction accompanying personal accomplishment while persons lower in *n* Achievement are more concerned with avoiding feelings of failure" (5, p. 386), i.e., "lower *n* Achievement scores imply relatively greater anxiety about failure" (5, p. 389).

Thus, recalling more of one's failures (interrupted tasks) under achievement-oriented conditions, the pattern of the *S* who scores high on *n* Achievement, is interpreted by Atkinson to express *S*'s "positive disposition to excel," by Alper to mean that the Weak Ego, overwhelmed by the experience of failure, protects self-esteem by taking the offensive; he admits his failures before others accuse him of having failed. Recalling more of one's successes (completed tasks) under achievement-oriented conditions, the pattern

of the low *n* Achiever, is interpreted by Atkinson (5) as an expression of anxiety over failure and by Alper as a way of supporting self-esteem through focusing on tasks one did complete. Seemingly opposite defense patterns, in other words, are seen to be operating to produce the same recall patterns.

The results in Part II of this paper provide a basis for showing how the two interpretations may be brought into a common theoretical framework. By obtaining both Ego Strength and *n* Achievement measures on female college leaders, a group which according to previous studies should score high on *n* Achievement (10, 20), it was found that *Ss* whose self-ratings fit the Strong Ego pattern more often score *below* the group mean on *n* Achievement than above it. No relationship was found between *n* Achievement and the Strong Ego self-rating pattern in two control groups. Since, by definition, the leader group should value social acceptance, i.e., score high on *n* Affiliation, we now can relate the Atkinson (5) and Atkinson and Raphelson (6) findings to the Strong Ego selective recall patterns isolated by the writer. The greater recall of interrupted tasks under the Relaxed Orientation by *Ss* who score high on *n* Affiliation (6), and of completed rather than interrupted tasks under conditions which objectively stress achievement by *Ss* who score *low* on *n* Achievement (5, 6), is exactly the recall pattern expected of a Strong Ego *S*.

While the relationship between high *n* Achievement and the Weak Ego self-rating pattern could not be adequately tested in Part II because so few *Ss* in the leader group fitted the Weak Ego pattern, the inverse relationship between the Strong Ego self-rating pattern and *n* Achievement permits at least a partial reinterpretation of the role *n* Achievement plays in selective recall. The results of Part II do not support Atkinson's (5, p. 389) proposition that "lower *n* Achievement scores imply relatively greater anxiety about failure" if low *n* Achievement is combined with the personality structure of a Strong Ego *S*. On the contrary, these *Ss* are able to counteract the objective failure threat (the Strong Ego *Ss* in Part I obtained significantly more alternative solutions than did Weak Ego *Ss*). Counteraction suggests positive, constructive forces at work rather

than a mechanism of "anxiety-avoidance." In other words, it may be the high  $n$  Achievers as well as Zeigarnik's (29) "ambitious" Ss, not the low  $n$  Achievers, who have the relatively greater anxiety about failure.

Further support for these interpretations comes from an analysis of the self-ratings of the leader group on the nine individual personality variables included in *The Psychological Insight Test*. As a group, these socially accepted Ss significantly more often rated themselves high rather than low, i.e., above the average for college women, not below the average, on  $n$  Counteractive Achievement, Conative Con-junctivity and Ego Strength, and low rather than high on  $n$  Recognition, Dejection-Pessimism and Ego-Ideal Intrgression. The variables on which they score equally often at the high or the low end are Narcism,  $n$  Defend-ance, and  $n$  Dominance. In this group, the Ss accept as characteristic of themselves, then, only those variables which could be regarded as having a positive value in our culture. This finding is consistent with Martire's (17) finding that low  $n$  Achievers are less dissatisfied with themselves, less in "self-conflict," than are high  $n$  Achievers.

The low  $n$  Achiever who is a Strong Ego, in other words, is low on achievement motivation as long as the *objective* situation is unrelated to achievement. His Achievement needs are *not constantly* under tension. Yet when achievement is realistically called for, and failure is a realistic possibility, the Strong Ego can and does counteract it. The counteraction shows itself both in his selective recall and in his performance data. It should not be surprising, therefore, to find significant differences in selective recall between high and low  $n$  Achievers consistently reported only for Achievement Orientation (5, 6), while performance studies of high and low  $n$  Achievers continue to yield inconsistent results (6, 7, 12, 27). In some of these studies, high  $n$  Achievers perform better than low  $n$  Achievers under both Task Orientation and Achievement Orientation, in others only under Achievement Orientation. To understand these results better, it may be necessary to take into account both  $n$  Achievement and  $n$  Counteractive Achievement, to differentiate, for example, between high  $n$  Achievers who are high on  $n$  Counteractive Achievement and high  $n$

Achievers who are low on  $n$  Counteractive Achievement. Moreover, while low  $n$  Achievement and high  $n$  Counteractive Achievement appears to be the Strong Ego pattern, theoretically, Weak Ego Ss could fall either into a low  $n$  Achievement, low  $n$  Counteractive Achievement or high  $n$  Achievement, low  $n$  Counteractive Achievement group. The two Weak Ego Ss in the present leader group, for example, would fit the first pattern since they rated themselves low on  $n$  Counteractive Achievement and scored low on  $n$  Achievement as measured by the TAT technique.

While the relationships between low  $n$  Achievement and Strong Ego self-ratings hold statistically, and thus warrant the reinterpretation of the selective recall findings of studies done within the  $n$  Achievement framework, certain problems in the data of the present study, and certain objections which can be raised concerning them, need to be considered. The first problem centers around a difficulty raised in Part I. The recall patterns of Strong and Weak Ego Ss were not statistically different under Task Orientation, though the overall differences between the sessions were significant. The earlier study (2) had indicated that both orientations would yield significant recall differences within sessions. The absence of significant recall differences under Task Orientation is not peculiar to the present study (see 3, 6, 13, 26). While no general explanation is available, a factor that may have been important in the present study may also have affected other studies. In the experiment reported in Part I, the experimental atmosphere may not have been as nonthreatening to self-esteem as *E* had intended, even though the instructions were the same as those previously used. In the earlier study (2), the Ss were familiar with the clinic setting and with the clinic personnel. They came to the clinic many times since all were participating in an extensive personality study. In the present experiment, the Ss served only in the one experiment. If this explanation has validity, it suggests that at a given point in the ego-stress continuum, ego tensions begin to function for Strong Ego Ss also, and at this point there are no differences in the recall patterns of the two groups of Ss. At a further point along the continuum, the recall patterns would again differ.



A second problem in relating the recall patterns emerging in the *n* Achievement studies to Alper's patterns, stems from possible differences in an "Achievement" vs. an "ego oriented" or "self-esteem threatening" atmosphere. While the terms themselves do not necessarily indicate important motivational differences, the writer would agree that so long as the conditions used to arouse the motive center around achieving, i.e., making a good record for oneself against some objective standard of excellence, the term "Achievement Orientation" expresses very precisely the motivation the *E* intends to have aroused. Yet in both the Atkinson and in the Alper studies, the arousal of more than achievement needs seems to have been intended since in both, the *S* is prevented from doing well on (i.e., completing) "tests." In the Atkinson study, *S* was told that the tests were "developed by psychologists during the war years in order to select people with high executive capacity, intellectual alertness, the capacity for making quick decisions, and leadership" (5, p. 382). In Alper's study, the tests were presented as "a brief intelligence test which the Army was finding useful in selecting candidates for Officer Training School" (2, p. 106). In other words, in both studies it was intended that *S* experience failure and that self-esteem be threatened. It is quite possible, of course, that Alper's instructions would be more threatening to draft age males than Atkinson's, if only because the incompleteness occurred under more public conditions in the one than in the other; in Alper's study, an accomplice, of college age, "succeeded" on tasks which *S* was experimentally doomed to fail. If these differences affect recall, however, theoretically they should sharpen rather than alter the direction of recall differences.

A third problem arises with respect to whether the correlation between self-rating scores and *n* Achievement scores (the procedure followed in Part II) should be taken at face value since recent studies of the relationship between self-rating scores on achievement (*v* Achievement) and TAT measures of achievement motivation (*n* Achievement) are not entirely consistent. Two studies (8, 9) have used male *Ss*, one (20) has used female *Ss*. For male *Ss*, de Charms *et al.* (9) report a positive correlation between *v* Achievement

and *n* Achievement, the *P* value they obtained being  $<.05$ . But Child *et al.* (8) found no relation between *v* Achievement and *n* Achievement in their study of college males. For a female sample of college leaders, Morrison (20) reports that *v* Achievement and *n* Achievement are not correlated, while in the present study of female college leaders, there was a significant *inverse* relation between self-rating scores and *n* Achievement for the Strong Ego *Ss* ( $P < .02$ ).

The only other data available for comparing the relationships between *v* Achievement and *n* Achievement center on the F-scale scores on males in the de Charms *et al.* study, and on females in the present study. For males, de Charms *et al.* (9, p. 416) report that *v* Achievement correlates positively with total score on the F scale, while *n* Achievement does not. In the present study, *n* Achievement was also not related to F-scale scores, but self-rating scores were: Strong Ego *Ss* tended to score *below* the group mean on the F scale ( $P < .10$ ), while *Ss* who did not fit the Strong Ego pattern, treated as a group, significantly scored *above* the group mean on the F scale ( $P < .05$ ). In other words, it is the non-Strong Ego *Ss* who are the more conformist, the more rigid *Ss* among these college leaders, including the two *Ss* whose self-rating scores fitted the Weak Ego pattern. The interpretation in Part I, that Strong Ego *Ss* were less rigid than Weak Ego *Ss*, as measured by their ability to obtain more alternative solutions in the scrambled phrases task, is consistent with the F scale findings in Part II and with the data in the Child *et al.* (8) study.

Since many variables have been shown to influence the strength of the *n* Achievement variable, including the factor of the sex of the picture stimulus, further experimentation is necessary before the differences in findings of various studies can be clearly understood. The specific pictures used, the experimental atmosphere in which *n* Achievement is measured, even the questionnaire items<sup>21</sup> used in the self-rating measurements have varied greatly

<sup>21</sup> In all studies, the items have been derived from the Murray *et al.* (21) approach to measuring need strength by the questionnaire method. The length of the scales and the specific items covered have differed, however.

from one study to another. On the other hand, the relationships among these variables revealed in the present study are consistent with the writer's interpretation that Strong Ego Ss score *low* on *n* Achievement essentially because their achievement needs are *not* under tension rather than because of concern "with avoiding feelings of failure" (5, p. 386). This interpretation is an essential link for relating the selective recall pattern of the low *n* Achiever in the Atkinson (5) and Atkinson and Raphelson (6) studies to Alper's Strong Ego pattern and thus for reaffirming both the theoretical framework presented earlier by the writer (2) and its applicability to previously reported contradictory or equivocal findings of other studies (3).

#### SUMMARY

Two experiments are reported here. The first was designed to validate the two major recall patterns that had emerged in a previous study of the relationship between personality structure and the recall of interrupted vs. completed tasks, the Strong Ego pattern and the Weak Ego pattern (2). For this purpose, two groups of college males, a Strong Ego and a Weak Ego group, were selected in advance of the experiment by means of a specially constructed self-rating scale and a clinical interview. Otherwise, the design of the experiment, procedure, and materials were the same as in the previous study. Using only the total recall score pattern as a measure of the change in direction of recall from a task-oriented to an ego-oriented session, it was possible to predict correctly the group membership of these Ss, as Strong or as Weak Ego Ss, significantly beyond chance. These results confirm, therefore, the earlier finding that Strong Ego Ss recall significantly more completed tasks as the experimental conditions become more threatening to self-esteem; Weak Ego Ss, interrupted tasks. In the present study, the recall patterns of the two groups did not differ significantly within the task-oriented session, but did differ within the ego-oriented session.

The second experiment was designed to test the relationships between these recall patterns and the *n* Achievement findings reported by Atkinson (5) and Atkinson and Raphelson (6). These authors have reported that high *n* Achievers, as measured by the TAT method of

McClelland *et al.* (18), recall more interrupted than completed tasks as the experimental conditions become more clearly achievement-oriented, while low *n* Achievers show the opposite recall pattern.

Three groups of female Ss were used, and both *n* Achievement and Ego Strength were measured. One group was made up only of college leaders who had been elected to high office by their peers. Following Morrison (20), these Ss should be a high *n* Achievement scoring group and could serve, therefore, as the experimental group. Two other nonleader groups served as controls. A significant *inverse* relationship between Strong Ego self-rating scores and *n* Achievement was obtained only for the leader group. The two measures were not correlated in the control groups. These findings are used to show that the recall pattern of the low *n* Achiever, as reported by Atkinson (5) and Atkinson and Raphelson (6), corresponds to Alper's Strong Ego pattern. The usual interpretation of a low *n* Achievement score as implying "anxiety over failure" is re-examined in the light of these findings.

It cannot be concluded from these data, however, that the selective recall pattern of a high *n* Achiever would necessarily fit Alper's Weak Ego pattern. Not only were there too few Weak Ego Ss in the college leader sample to test the relationship between Ego weakness and *n* Achievement, but also, and of more importance from the point of view of theory, Ego Strength and Ego Weakness, as measured in this study, are not at opposite ends of a single personality continuum. We cannot assume, therefore, that Ego Strength and Ego Weakness would function with respect to *n* Achievement in diametrically opposite ways. Yet given a high *n* Achiever whose self-ratings are congruent with the Weak Ego personality pattern, including low *n* Counteractive Achievement, it would be consistent with the results of these experiments to expect him to show the selective recall and performance patterns characteristic of Weak Ego Ss.

The theoretical framework presented earlier by the writer (2) is thus reaffirmed. The usefulness of this framework for reconciling seemingly contradictory selective recall and performance data reported in other studies, as previously discussed (3), is also reaffirmed.

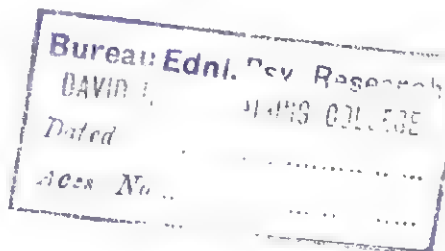


Specific findings concerning *n* Achievement in females, including the importance of the sex of the picture stimulus used and the experimental atmosphere in which *n* Achievement is measured, are discussed.

## REFERENCES

1. ALPER, THELMA G. Memory for completed and incompleting tasks as a function of personality: an analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403-420.
2. ALPER, THELMA G. Memory for completed and incompleting tasks as a function of personality: correlation between experimental and personality data. *J. Pers.*, 1948, **17**, 104-137.
3. ALPER, THELMA G. The interrupted task method in studies of selective recall: a reevaluation of some recent experiments. *Psychol. Rev.*, 1952, **59**, 71-88.
4. ALPER, THELMA G., & BLACK, ADELAIDE A. The effect of instruction, task and population sample on mental set. *Amer. J. Psychol.*, 1949, **62**, 295-299.
5. ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381-390.
6. ATKINSON, J. W., & RAPELSON, A. C. Individual differences in motivation and behavior in particular situations. *J. Pers.*, 1956, **24**, 349-363.
7. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469-476.
8. CHILD, IRVIN L., FRANK, KITTY F., & STORM, T. Self-ratings and TAT: their relations to each other and to childhood background. *J. Pers.*, 1956, **25**, 96-114.
9. DECHARMS, R., MORRISON, H. W., REITMAN, W., & MCCLELLAND, D. C. (Eds.) *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. Pp. 414-427.
10. FIELD, W. F. The effects on thematic apperception of certain experimentally aroused needs. Unpublished doctor's dissertation, Univer. of Maryland, 1951.
11. FISHER, R. A. *Statistical methods for research workers*. (6th ed.) Edinburgh: Oliver and Boyd, 1936.
12. FRENCH, ELIZABETH G. Some characteristics of achievement motivations. *J. exp. Psychol.*, 1955, **60**, 232-236.
13. GLIXMAN, A. F. An analysis of the use of the interruption-technique in experimental studies of repression. *Psychol. Bull.*, 1948, **45**, 491-506.
14. HORN, D. A study of personality syndromes. *Charact. and Pers.*, 1944, **12**, 257-274.
15. LEWIS, HELEN B. An experimental study of the role of the ego in work. I. The role of the ego in co-operative work. *J. exp. Psychol.*, 1944, **34**, 113-126.
16. MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, 1947, **18**, 50-60.
17. MARTIRE, J. G. Relationships between self-concepts and differences in strength and generality of achievement motivation. *J. Pers.*, 1956, **24**, 364-375.
18. MCCLELLAND, D., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
19. The Michigan Picture Test. Chicago, Ill.: Science Research Associates, 1953.
20. MORRISON, H. W. Validity and behavioral correlates of female need for achievement. Unpublished master's thesis, Wesleyan Univer., 1954. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. P. 422.
21. MURRAY, H. A. *et al. Explorations in personality*. New York: Oxford Univer. Press, 1938.
22. MURRAY, H. A. *Thematic Apperception Test*. Cambridge, Mass.: Harvard Univer. Press, 1943.
23. ROSENZWEIG, S. An experimental study of "repression" with special reference to need-persistent and ego-defensive reactions. *J. exp. Psychol.*, 1943, **32**, 64-94.
24. ROSENZWEIG, S., & MASON, GWENDOLYN. An experimental study of memory in relation to the theory of repression. *Brit. J. Psychol.*, 1934, **24**, 247-265.
25. SANFORD, N., & RISSER, J. What are the conditions of defensive forgetting? *J. Pers.*, 1948, **17**, 244-260.
26. SEARS, R. R. Personality. In C. P. Stone, & D. W. Taylor (Eds.), *Annual review of psychology*. Stanford: Annual Reviews, 1950, Vol. I.
27. SNEDECOR, G. W. *Statistical methods*. (4th ed.) Ames, Iowa: Iowa State College Press, 1946.
28. VEROFF, J., WILCOX, SUE, & ATKINSON, J. W. The achievement motive in high school and college age women. *J. abnorm. soc. Psychol.*, 1953, **48**, 108-119.
29. ZEIGARNIK, BLUMA. On finished and unfinished tasks. In W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Harcourt Brace, 1938. Pp. 300-314.

Received August 20, 1956.



# THE RELATIONSHIP BETWEEN ADVANCED AGE AND MENTAL ABILITIES<sup>1</sup>

CHARLES R. STROTHER, K. WARNER SCHAE, AND PAUL HORST

*University of Washington*

**I**N RESEARCH on aging, longitudinal studies in which the same population is followed over a long span of years are obviously to be preferred. There are an increasing number of such studies covering the periods of childhood, adolescence, early and middle adulthood (1, 2, 5), but a paucity of longitudinal data is as yet available on older age groups. The principal difficulties involved in such studies are the long period of time required for completion and the unavoidable attrition as the original subjects advance in age, due to such factors as mobility, illness, and death.

Cross-sectional studies, on the other hand, in which population samples in different age groups are compared, are open to several criticisms (3, 10). One criticism is concerned with the differences in educational background that are likely to characterize different age groups, usually to the advantage of the younger group. A second criticism is directed toward differences between age groups with respect to range of talent. For various reasons, it is difficult to obtain comparably broad samples of older and younger populations. Older subjects are more frequently drawn from institutions for the aged or from small groups of volunteers, while the younger subjects represent broader segments of the population. There are also differences in motivation and in attitudes toward, and familiarity with, psychological tests, which may tend to favor the younger groups. Finally, while the disabilities of older people with respect to speeded tests have been well established, few attempts have been made to apply appropriate corrections or to employ tests where speed is not a factor.

While these difficulties are serious, they may be minimized by restricting the groups compared to a fairly narrow age range. In the present study, the age range has been limited to the period 70-88 years. This restriction

<sup>1</sup> This investigation was part of an interdisciplinary research project at the University of Washington, supported by a grant from the Agnes H. Anderson fund. It was reported at the 1956 meeting of the American Psychological Association. The authors wish to express their appreciation to Mary Marjorie Campbell who assisted in collection of the data.

made possible the selection of subjects who were relatively homogeneous with respect to educational background, socioeconomic status, motivation, and familiarity with psychological tests. In this restricted population, differences in "primary mental abilities," "memory," "psycho-motor speed," and "motor rigidity" were determined by appropriate tests. Correction factors were then applied to the mental abilities scores to control the effects of loss of memory, reduced speed, and increased rigidity.

## METHOD

### *Subjects*

All subjects (Ss) were volunteers obtained from the membership of a retired university faculty group or from among retired academic or professional workers who had responded to appeals in newspapers published in the city in which the university was located. The Ss were interviewed before being accepted for the research project, at which time it was explained that about eight hours, spread over several days, would be required for the physical examinations, psychological tests, and interviews involved in the larger project of which the present study is a part. At this point, the S was given an opportunity to withdraw from the study if he wished. It seems reasonable to conclude that the Ss who participated were at least as highly motivated as the average volunteer group.

The group consisted of 25 male and 25 female college graduates, ranging in age from 70 to 88 years, with a mean age of 76.5 years. Clinical and laboratory examinations indicated that all members of the group were in fair to superior physical condition for their age. No psychiatric complaints or symptoms were noted by the examining physicians.

### *Tests*

In order to obtain estimates on more specific aspects of mental ability than would be yielded by a global score, Thurstone's Primary Mental Abilities Test (Intermediate Form) (12), which provides separate scores for Verbal-meaning, Space, Reasoning, Number, and Word-fluency factors, was selected. Previous work (6) had shown that the reliability and the range of difficulty for older subjects were adequate.

Since age changes in speed, memory, and rigidity are known to occur and might be expected to have some relationship to performance on the Primary Mental Abilities Test, measures of these variables were obtained. The Wechsler Memory Test (13) was selected as a measure of memory loss and Schae's Test of Behavioral Rigidity (9) was used to obtain measures of "psycho-motor speed" and "motor rigidity."

All tests were administered as individual tests and



the order of administration was randomized in order to avoid systematic effects of fatigue.

### RESULTS

The subjects were rank ordered by chronological age and divided into five groups of five males and five females each. A comparison by  $\chi^2$  showed no significant differences in years of education among the five groups. Standard Primary Mental Ability (PMA) scores with means of 50 and sigmas of 10 were then computed, by the use of Thurstone's conversion chart, from the norms for his 17-year-old group. Comparable standard scores for the speed and rigidity measures were computed, using Schaie's tables for his younger groups, ages 18-26. Standard scores on the Memory scale were computed from Wechsler's norms for his 20-30-year-old group (13).

Table 1 gives the mean standard scores for each of the five groups on all test variables. Examination of this table indicates that, for the 71-year-old group, the Reasoning and Spatial scores are below the mean of the 17-year-old reference population. If it can be assumed, on the basis of the amount of education and of previous occupational level, that the present group at an earlier age would have been above the mean of the general population on these tests, then an unspecified amount of loss in reasoning and spatial abilities has occurred prior to age 71. The Verbal-meaning, Number, and IQ scores are somewhat above the means for the younger reference population. Whether any loss has occurred on the Verbal and Number factors cannot be determined from these data. The relatively low scores of the 71-year-olds on the Word-fluency factor is most

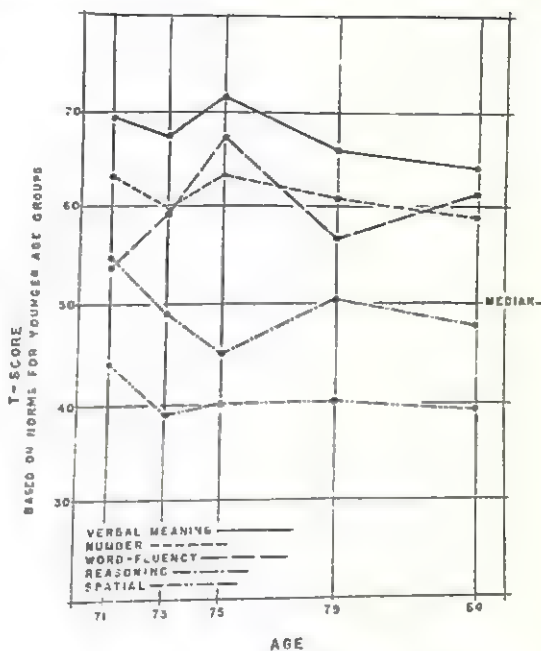


FIG. 1. MEAN T SCORES ON PRIMARY MENTAL ABILITIES FOR SUCCESSIVE AGE GROUPS

probably an artifact of sampling. The memory, speed, and motor-rigidity scores fall below the means for their respective reference groups.

Figure 1 presents graphically the subgroup means for the PMA scores. For the Spatial and Reasoning factors, the scores show a steady decline after age 71. The Verbal-meaning, Number, and Word-fluency scores remain fairly high in the 71-year-old group, and then decline.

Figure 2 shows the curves of decline for the speed, memory, and motor-rigidity factors. There is little difference between the scores for the 71- and 73-year-old groups, but beyond age 73 the curves drop steadily. The most rapid loss occurs in memory.

Statistical analysis, using the *t* test to evaluate the significance of the difference between the youngest and the oldest groups, shows a decline significant at the .05 level on all factors except for Word-fluency.

The next step in the analysis was to determine whether the decline in PMA scores would be affected if corrections were made for the effects of lowered speed, increased rigidity, and memory loss. The previously described independent measures of these abilities were therefore correlated with the PMA scores. Intercorrelations among all these measures were computed on the IBM Type 604 electronic calculating punch, using the standard

TABLE 1  
MEAN T SCORES FOR SUBGROUPS  
(Young Adult Norms,  $M = 50$ ,  $\sigma = 10$ )

Group	Mean Age	V*	S	R	N	W	IQ	Sp.	Mem	M-Rig
1	71	58.5	41.9	46.4	57.9	47.6	54.3	44.2	39.4	46.4
2	73	57.1	37.6	44.1	55.6	54.0	51.6	44.2	43.6	46.8
3	75	59.4	38.2	40.6	58.1	60.1	54.1	43.8	41.8	45.7
4	79	52.8	35.9	39.8	53.4	48.3	45.4	41.6	34.1	41.0
5	84	45.8	33.5	32.0	48.0	48.3	40.5	33.2	29.0	37.9
$D_{t-s}$		12.7	8.4	14.4	9.9	-0.7	13.8	11.0	10.4	8.5

\* Key To Symbols:

V = Verbal Meaning  
S = Space  
R = Reasoning  
N = Number  
W = Word Fluency

IQ = Intelligence Quotient  
Sp = Speed  
Mem = Memory  
M-Rig = Motor Rigidity

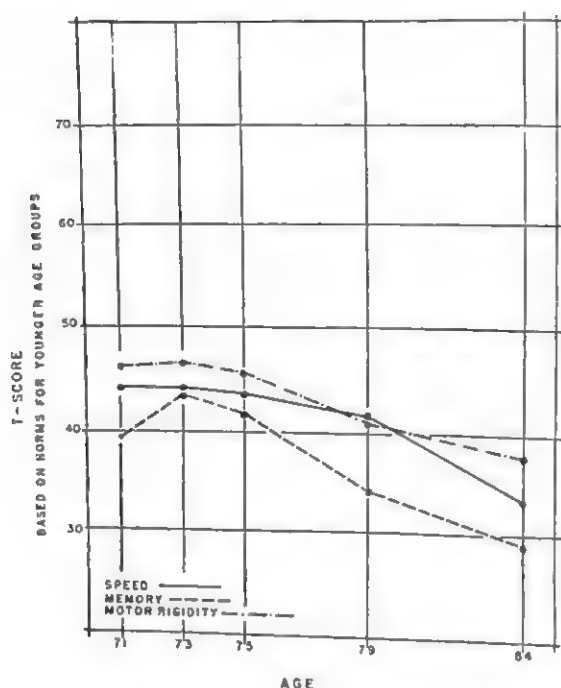


FIG. 2. MEAN T SCORES FOR MEMORY, SPEED AND MOTOR RIGIDITY

scores method (7). These intercorrelations, most of which are significant, are given in Table 2. Formulas were derived to correct for the effects of these factors on the PMA scores.<sup>2</sup> The correction formulas provide estimates of those parts of the PMA scores which are independent of the combined influence of the speed, rigidity, and memory scores, and yield corrected PMA measures in terms of *T* scores for the norm group.

It was not necessary to compute individual corrected PMA scores for the members of the experimental group in order to get estimates of the corrected subgroup means. These could be computed directly from the subgroup means given in Table 1 and the intercorrelations in Table 2, but the computations required the calculation of the beta regression vector for each PMA variable on the correction variables and also its multiple correlation with them.

<sup>2</sup> Materials presenting the mathematical derivation of the correction formulas, computational instructions, and tables of intermediate computations have been deposited with the American Documentation Institute. Order Document No. 5299 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 2  
TEST INTERCORRELATIONS

	S	R	N	W	IQ	Sp.	Mem.	M-Rig.
Verbal Meaning	.348	.612	.300	.172	.630	.676	.375	.501
Space		.361	.193	.093	.470	.145	.266	.263
Reasoning			.395	.354	.757	.674	.571	.462
Number				.287	.784	.411	.318	.325
Word-fluency					.570	.396	.452	.025
Composite-IQ						.667	.554	.445
Speed							.410	.318
Memory								.406
Motor-Rigidity								

TABLE 3  
BETA REGRESSION WEIGHTS AND MULTIPLE CORRELATIONS

	V	S	R	N	W	I.Q.
Speed	.571	.010	.495	.305	.298	.497
Memory	.013	.187	.292	.120	.429	.280
Motor Rigidity	.315	.184	.186	.180	-.244	.174
$\Sigma$	.899	.381	.973	.605	.483	.951
$R^2$	.548686	.099584	.586294	.222015	.305816	.564049
$R$	.741	.316	.766	.471	.553	.751

TABLE 4  
MEAN T SCORES AFTER CORRECTIONS FOR THE EFFECT OF SPEED, MEMORY LOSS AND INCREASED MOTOR RIGIDITY

Mean Age	V	S	R	N	W	I.Q.
71	69.5	44.4	54.7	63.1	53.6	66.3
73	67.2	38.9	49.1	59.9	59.2	60.4
75	71.5	40.1	45.1	63.3	67.4	65.5
79	65.9	40.2	50.4	60.7	56.6	58.4
84	64.1	39.3	47.9	58.9	61.2	60.4
$D_{1-8}$	5.4	5.1	6.8	4.2	-7.6	5.9

The regression weights and multiple correlations are given in Table 3, and the corrected subgroup means in Table 4. Strictly speaking, the correlations from which the regression weights and multiple correlations are derived should be based on a single norm group. Since this was not possible, it was necessary to use the correlations in Table 2, based on the experimental group.

Even though it was not necessary to use the individual scores to calculate the corrected subgroup means for the experimental group, corrected PMA scores may be computed for any given individual by means of Table 5.<sup>3</sup> Assume that a person's corrected ( $P_c$ ) score

<sup>3</sup> The derivation of this table is indicated in Section C of the material deposited with the American Documentation Institute. See Footnote 2.



TABLE 5  
WEIGHTS USED IN CALCULATING INDIVIDUAL  
CORRECTED PMA SCORES

	V	S	R	N	W	I.Q.
$W_P$	1.489	1.054	1.555	1.134	1.200	1.514
$W_B$	.850	.011	.770	.346	.358	.752
$W_M$	.019	.197	.454	.136	.515	.424
$W_R$	.469	.194	.289	.204	.293	.263
F	42.5	17.4	47.9	27.6	19.0	46.3

is desired for any given PMA score. Let his  $T$  scores on speed, memory, and motor rigidity be respectively  $C_S$ ,  $C_M$ , and  $C_R$ . His  $P_o$  score is given by

$$P_o = PW_P - C_S W_B - C_M W_M - C_R W_R + F.$$

The values for the  $W$ 's and for  $F$  are taken from the appropriate column of Table 5. For example, suppose a person's  $P_o$  score for the verbal factor is desired and that his verbal score is 47, and his speed, motor, and rigidity scores are, respectively, 53, 49, and 45. Using the values in the first column of Table 5 we have

$$P_o = 47(1.489) - 53(.850) - 49(.019) - 45(.469) + 42.5 = 45.4.$$

Curves for the corrected subgroup means are shown in Figure 3. When corrections are made

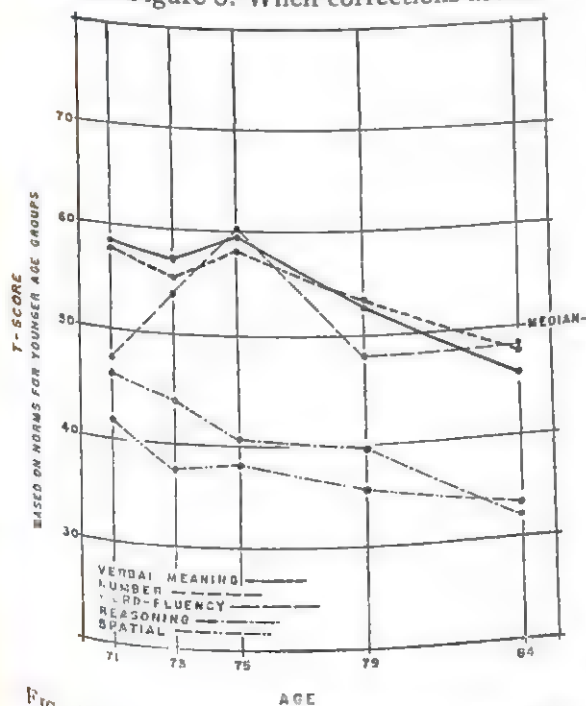


FIG. 3. MEAN  $T$  SCORES ON PRIMARY MENTAL ABILITIES CORRECTED FOR LOSS OF SPEED, MEMORY, AND FOR INCREASED MOTOR RIGIDITY

TABLE 6  
STANDARD SCORES ON THE PRIMARY MENTAL  
ABILITIES TEST FOR SUBJECTS OVER  
EIGHTY YEARS OF AGE

Age	Verbal-meaning	Spatial	Reasoning	Number	Word-fluency
Females:					
80	58	33	39	39	40
80	56	32	48	56	49
80	56	39	42	49	38
81	42	36	30	50	58
82	54	35	33	58	39
82	34	27	27	27	40
83	50	27	27	52	41
88	35	37	34	35	67
Males:					
80	46	33	32	49	52
81	49	37	36	56	47
82	51	40	33	41	52
82	51	38	33	60	37
83	50	41	35	52	65
84	50	34	30	50	42
85	50	27	44	61	40
88	42	33	27	35	54

for losses in speed and memory and for an increase in motor-rigidity, the age curves for the Primary Mental Abilities flatten out appreciably. The Verbal, Number, and Word-fluency scores remain approximately 1  $SD$  above the mean of the younger age groups at age 84. Scores on the Spatial factor, while significantly below the mean for the younger ages, show no appreciable change over the age range, 73-84.

The range of differences in Primary Mental Abilities scores for individuals of advanced age is of interest. Table 6 shows the uncorrected standard scores (with a mean of 50 and an  $SD$  of 10) on this test for the oldest subjects in the present group. Of the total of 16 in this age range, the verbal meaning scores of ten  $S$ s and the numerical ability scores of nine remain at or above the mean for an unselected population of 17-year-olds.

## DISCUSSION

The retention of verbal ability shown here is consistent with the results of other studies. Owen's data (3) and those of Bailey and Oden, (2), both obtained by longitudinal methods, indicate that some verbal abilities continue to increase up to age 50. Doppelt and Wallace (4) have recently reported normative data for the Wechsler Adult Intelligence Scale on a fairly large and representative sample of older persons. The Vocabulary, Arithmetic, and Information tests show the smallest decline up

to age 70. For their oldest age group, with a mid-point age of 79.5, the Vocabulary score is only approximately .5 *SDs* below the mean of the reference population.

The studies most directly comparable with the present one are those of Schaie, Rosenthal, and Perlman (6), and of Sward (11). In the Schaie study, the oldest age group ranged from 71–78 years and the general educational level was considerably below that of the present group. On the *PMA* test, Schaie's older *Ss* fell below the mean of the 17-year-old reference population on all factors. The lowest scores were obtained on the Space and Reasoning factors, as in the present study, and the highest scores were on Number and Verbal-meaning, as in the present case. However, the rank order of magnitude of these scores was different in the two studies. The earlier study showed the greatest loss on the Reasoning factor and the least on Number, while the present study showed the greatest loss on the Spatial factor and the least on Verbal-meaning.

Sward's study is of particular interest because his *Ss* were drawn from university faculties and thus were fairly closely comparable to those used in the present study. His older group ranged in age from 60–79, with a mean CA of 66.2; his younger group ranged from 21–42, with a mean CA of 31.4. Unfortunately, from the standpoint of any direct comparison of his results with ours, Sward used a special test battery, the correlations of which with *PMA* scores are unknown. However, on all except two of his tests the older group did significantly less well than the younger. On his "word-meaning" test, the older group did as well as the younger. On a "synonym-antonym" test, the older *Ss* were significantly better and on an "index of verbalization" which Sward constructed, the older *Ss* also scored higher. These results are consistent with those obtained in the present study and lend support to the conclusion that, at least in a population of superior intelligence, verbal abilities remain at a high level into the middle seventies.

#### SUMMARY

Scores on primary mental abilities, speed, memory, and motor-rigidity were obtained for

a group of fifty college graduates, ranging in age from 70–84 years. Differential decline in these abilities with advancing age is shown, with the earliest and largest losses occurring in memory, in speed, and in reasoning and spatial abilities. Word-fluency, verbal-meaning, and numerical abilities in this group of superior individuals remain well above the mean for young adults until the middle seventies. Verbal-meaning and numerical ability scores remain at or above the mean for the 17-year-old reference population in a majority of the *Ss* over eighty years of age. Since this study is cross-sectional, rather than longitudinal, the possibility that selective factors may account for the high scores of the older *Ss* cannot be ruled out.

#### REFERENCES

1. BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychol.*, 1955, 10, 805–818.
2. BAYLEY, NANCY, & ODEN, M. H. The maintenance of intellectual ability in gifted adults. *J. Gerontol.*, 1955, 10, 91–107.
3. DENNIS, W. *Age and behavior*. Randolph Field, Texas: U. S. School of Aviation Medicine, 1953.
4. DOPPELT, J. E., & WALLACE, W. L. Standardization of the Wechsler Adult Intelligence Scale for Older Persons. *J. abnorm. soc. Psychol.*, 1955, 51, 312–330.
5. OWENS, W. A. Age and mental abilities: a longitudinal study. *Genet. Psychol. Monogr.*, 1953, 48, 3–54.
6. SCHAE, K. W., ROSENTHAL, F., & PERLMAN, R. M. Differential mental deterioration of factorially "pure" functions in later maturity. *J. Gerontol.*, 1953, 8, 191–196.
7. SCHAE, K. W. The standard score method of computing intercorrelation matrices on the IBM type 604 electronic calculating punch. *IBM Technical Newsletter*, New York, 1955, No. 12.
8. SCHAE, K. W. *Examiner manual for the test of behavioral rigidity*. American Documentation Institute, Library of Congress, Washington, D. C., Document No. 4651.
9. SCHAE, K. W. A test of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1955, 51, 604–610.
10. SHOCK, N. W. Gerontology. *Ann. Rev. Psychol.*, 1951, 2, 353–370.
11. SWARD, K. Age and mental ability in superior men. *Am. J. Psychol.*, 1945, 58, 443–479.
12. THURSTONE, L. L., & THURSTONE, T. G. *Primary Mental Abilities Test*. Chicago: Science Research Associates, 1947.
13. WECHSLER, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, 19, 87–95.

Received August 13, 1956.



# JUDGMENTS OF CONSENSUS DURING GROUP DISCUSSION<sup>1</sup>

PHIL STONE AND JOE KAMIYA

*University of Chicago*

IF THE members of a group are asked to record privately their agreement or disagreement with a given proposition and to estimate the proportion of the group that shares the same opinion, the resulting data can be analyzed to reveal the degree of accuracy in the members' appraisal of their group's consensus. In several investigations (1, 2, 3, 4, 5) this technique, or some variant of it, has been used on a wide variety of opinion and attitude areas. Such studies have generally shown that individuals can estimate the consensus of their group with some degree of accuracy. Presumably, the estimates are based at least partly on knowledge acquired from past contact with whatever cues to the group's opinions that were available. The present study, with the general aim of shedding light on the manner in which the cues influence the estimates, undertakes a descriptive preliminary survey of judgments of group opinion in small discussion groups, where cues for the estimation of group opinion are provided through a discussion of the issues by the members themselves.

The questions explored are: Is the degree of estimation accuracy achieved through group discussion markedly different from that found in previous investigations where conditions of group discussion were not used? What is the nature of the errors? How are changes in estimates and in opinions related to the discussion content? Is accuracy increased by discussion?

## METHOD

### *Subjects*

The subjects (Ss) were six groups of high school students in Evanston, Illinois. The *N* of each group varied from 14 to 16. Both sexes were approximately equally represented in all classes. Ninety-two per cent of the students were either 15 or 16 years old, the remainder being 17.

### *Topics of Discussion*

In order to establish the discussion itself as the primary source of cues for the estimates of group consensus,

<sup>1</sup> This study is part of a project supported by a grant from the Social Science Research Council.

the topics presented were novel to the group. Pretests of comparable groups showed that the topics had probably not been discussed by the Ss before the study, but were of sufficient interest to maintain active discussion. In their order of presentation, the topics were: (a) whether the state law should be changed so that some high school juniors and seniors could drive the school buses at their school, (b) whether bicycle riders should be required by law to walk their bicycles across all intersections, and (c) whether women (adults) would be suitable as school bus drivers.

### *Procedure*

The Ss were seated facing each other around a circle. Response forms were provided each S. The experimenter (E) began the session for each group by saying that the purpose of the study was to survey their opinions on some topics through group discussion. The Ss were told to try to reach a group decision on the issues presented them.

The E explained that during the discussion he would occasionally interrupt the discussion, and request each person to record on the form his opinion on the specific proposition then being discussed. Only "yes" or "no" opinions were permitted, corresponding to agreement or disagreement with the proposition. Immediately after indicating his own opinion, S was required to record his estimate of the number of persons in the group who held the same opinion as he did. Each recording of an opinion and corresponding estimate is hereafter referred to as an "item."

The first item on each main topic was administered before any discussion occurred. The S was asked to give a favorable or unfavorable (yes or no) opinion and a consensus estimate on the topic as worded above. Subsequent items dealt with various facets of the main topic, as brought out by the discussants themselves. These items were careful formulations by E of the S's contributions into concrete propositions. A sample item from the first topic: "High school junior and senior bus drivers wouldn't be able to handle the rowdies on the bus." "Answer 'yes' if you agree; 'no', if you disagree, and estimate the number of people who marked the form the way you did."

The extent of discussion allowed between items was varied. At times only one S's contribution intervened; at the other extreme, as many as six different Ss were allowed to contribute before the next item was taken. Among all the items taken during the discussion of a main topic, some pairs of items were directly comparable, as a result of an S repeating a proposition which had been used as a previous item. These pairs were used to assess changes in the opinions and the estimates.

The Ss were informed in advance how long the discussion on each topic would last. At the end of the allotted period, E gave a final summary item and directed the discussion to the next main topic. The total time of discussion was about 45 minutes. The total

number of items taken in each group over all three major topics ranged from 18 to 26.

## RESULTS AND DISCUSSION

### *Accuracy of Estimation*

Presented first are measures of estimation accuracy over all items without regard to the nature of the discussion preceding the items. Two different measures were used. One measure was the difference between the estimated group opinion and the actual group opinion. The other measure was the Pearson product-moment correlation between estimated and actual opinion. These two measures were used in the analysis of the estimates of individuals as well as the pooled median estimates of the group.

*Accuracy of individuals.* For each *S*, the absolute difference between the estimated and actual group opinion was obtained on every item taken during the discussion period. The average of these differences over all items for a single *S* represented his average error score, in number of persons. The lowest average error obtained was 1.42 persons; the highest was 4.72 persons. The mean and *SD* of these average individual error scores are presented for each group in Table 1, Items 1 and 2.

For the correlational accuracy score of individuals over all items, *r*'s were computed between estimated and actual group opinion for each of 18 *S*s randomly selected from the six groups. The *r*'s ranged from .11 to .75 with a mean of about .40, values roughly comparable to the mean *r*'s of .69 and .44

reported by Travers (4) and Crow (2), respectively.

*Accuracy of median estimates.* For analysis of the estimates of the group as a whole, the median estimate was determined for each item. Since the task for each *S* was to estimate the number of *S*s who shared his opinion, a transformation was made before the median was computed to render the estimates of all *S*s into comparable estimates of the number of "yeses."

To measure group accuracy in terms of difference scores the absolute difference between the median estimate and the actual group opinion was determined for each item. The mean and *SD* of these error scores for each group are shown in Items 3 and 4 of Table 1. The mean error, expressed as a percentage of group size, is shown in Item 5. These values are also comparable to results found in other studies (3, 4).

As would be expected from the effects of pooling, the mean error of the estimates of the entire group is considerably smaller than the mean of the average error scores of single individuals shown in Item 1. However, further analysis indicates that the relative superiority of the group over the individual estimate is even greater than might be expected on these grounds: of the 88 *S*s tested, there are none with average error scores that were smaller than the mean error of the pooled estimates of their group.

In every group the correlation across all items between the median estimates of the entire group and the actual consensus was

TABLE 1  
SUMMARY OF RESULTS

Item	Group					
	I	II	III	IV	V	VI
Number of Subjects	14	15	14	14	16	15
Number of Items	19	18	23	26	19	20
1. Mean of individual average error scores	2.43	2.65	2.57	2.44	2.57	2.90
2. <i>SD</i> of individual average error scores	.41	.68	.56	.53	.80	.79
3. Mean error of group median	1.26	1.33	1.61	.96	1.42	2.10
4. <i>SD</i> of mean error	.84	1.00	1.00	.65	.90	1.56
5. Item 3 as percentage of group size	9%	9%	11%	7%	9%	14%
6. Correlation between group estimate and actual consensus	.95	.96	.90	.95	.97	.90
7. Mean algebraic error of factions	+2.12	+2.47	+2.21	+2.52	+2.89	+2.42
8. <i>SD</i> of algebraic error	2.66	3.43	3.13	2.96	4.27	3.80
9. Mean absolute error of factions	2.68	3.42	3.05	3.16	3.76	3.65
10. <i>SD</i> of absolute error	2.11	2.86	2.37	2.28	3.56	2.42
11. Correlation between group estimate and actual consensus, own opinion held constant	.67	.56	.64	.72	.57	.63



.90 or higher (see Item 6, Table 1). These  $r$ 's are similar to the  $r$ 's of .82 and .93 found by Campbell (1) and Travers (4).

*Accuracy and assumed similarity.* Previous studies (1, 4, 5) have shown that Ss' estimates of consensus are correlated with their own opinions. That is, those Ss who share a given opinion on an issue give a larger estimate of the numbers holding their opinion than do those who hold the opposite opinion. To determine the extent to which this tendency for Ss to assume similarity between themselves and others occurs under conditions of open discussion as created in this study, the median estimates of those whose own opinions were "yes" were computed separately from the median estimates of those whose opinions were "no." It was found that on every one of the 125 items in the entire study, the median estimate of one side was larger than the complement of the median estimate of the other side. The results of previous studies are thus clearly supported here.

In terms of accuracy, the foregoing finding means there is a general tendency for Ss on each side of an issue to overestimate the numbers who share their opinion. The average net tendency toward overestimation per item is shown in Item 7 of Table 1. These figures are averages of the *algebraic* differences between the median estimate of each side and the actual number of Ss on that side. A comparison of these differences with the absolute differences between the same measures (Item 9) clearly shows that the predominant error in each group is that of overestimation.

Consideration of the task confronting the Ss makes it clear that the direction and extent of error in estimation can not be uniform for all sizes of "factions" (those who share one opinion). If all group members hold the same opinion, the only error of estimation that can possibly occur is underestimation, since the highest estimate that can be given is limited to the total number of persons in the group. At the other extreme, if only one member holds a given opinion, he can err only in the direction of overestimation.

If such constraining effects of group size were the only factors influencing the errors in the estimates, equal but opposite errors in the median estimates of the two extremes of faction size would result. However, as

could be predicted from the observed widespread tendency to overestimate, such symmetry in the errors did not occur in the data. Analysis showed that factions comprised of about 10% of the group gave median estimates that averaged about 55-60% of the group size, while factions comprised of 90% of the group gave median estimates that were about 80% of the group size. The estimates of the intermediate sized factions fell approximately on a straight line between these two extremes, with greatest accuracy of estimation being found in factions comprised of approximately 75% of the group. Hence, it is clear that all but the largest factions tended to overestimate their size.

The fact that *all* factions, regardless of size, assumed on the average that more than half the group shared their own opinion (or that less than half of the group held the opposite opinion) accounts in part for the high degree of accuracy of the median estimate of the entire group, since a majority of the estimates on which the group median is based comes from the majority side, and the pooled estimate of the entire group is therefore partly determined by the actual consensus. The accuracy of the pooled median estimate is thus spuriously high when taken as an index of the judgmental achievement of the group. As an empirical fact, however, the high accuracy of median estimates should lend itself to the use of estimates for the indirect measurement of attitudes in small groups.

An attempt was made to determine whether any correlational accuracy would remain if the factor of assumed similarity were excluded. The S's own opinion was held constant by computing the accuracy correlations of the "yes" factions and the "no" factions separately across all items, and then averaging the two correlations. The results, shown in Item 11 of the table, indicate appreciable judgmental accuracy apart from the spurious effects of assumed similarity, with  $r$  ranging from .56 to .67 among the six groups.

#### *The Influence of Group Discussion on Estimates and Opinions*

The results to be reported focus on the relationship between the views expressed during the course of the discussion and the accompanying changes that occurred in the

estimates and opinions. All six groups were combined in this analysis. The changes were assessed by comparing the estimates and opinions taken on an item before discussion with those taken on a comparable item after discussion. From the six groups, a total of 49 comparable before-and-after item pairs was found.

Overall favorableness or unfavorableness toward the propositions presented of the views expressed in discussion was rated on a seven-point scale by two judges using verbatim transcripts. All relevant expressions of opinion intervening between the comparable items were rated as a single unit. The judges based their ratings on the number of different Ss speaking as well as the content of each S's speech. The inter-judge correlation over the 49 discussion units was .82. The average rating of the two judges on each item was transformed into a percentage according to its relative location within the range of all the ratings, in order to provide an index of the favorability of discussion that could be compared directly with the degree of favorableness of the group opinion measures.

*Changes in estimates.* Since the views expressed in support of a proposition (or in opposition to it) during discussion provide cues for the estimation of consensus, it is reasonable to expect the median estimate of the group to change in the direction of the expressed support, unless the expressed support is representative of the initial estimate. The expressed support was considered representative if it did not differ from the initial estimate by more than 10% of group size. Eliminating these representative cases, the median estimate shifted toward the level of expressed support in 35 out of 41 cases ( $p < .001$ ).<sup>2</sup>

*Changes in actual opinion.* Since the views

<sup>2</sup> For assessing the significance of this proportion, the value for the chance proportion under the null hypothesis was determined empirically, following the method used by Crow (2) for a somewhat different problem. The relative frequency with which the obtained changes occur in the expected direction was determined, using level of expressed support in every other discussion unit in the study as a base of prediction. This gave a total of  $49 \times 48$ , or 2352, comparisons. The proportion of cases in which the estimates shifted in the direction of these other ratings of discussion units was .57. This value was taken as  $p$  in the binomial distribution, against which the obtained proportion of 35/41 was evaluated.

expressed during discussion may affect opinions as well as estimates, a parallel analysis was undertaken of the changes in the opinions. The results were again clear cut. In 32 out of 41 cases, the actual group opinion shifted in the direction of the expressed support. This obtained proportion, evaluated against its chance level of .53 (obtained by the same procedure as for the estimates) is significant at the .001 level.

*Other findings.* The correlation between changes in the median estimates of the group and the changes in actual group consensus was found to be .89, significant at the .001 level.

It might be expected that the estimates would shift more in the direction of the expressed support than would the opinions, on the ground that the Ss are likely to think others are more affected by what is said than they are themselves. However, the results from several methods of analysis point to no significant difference between the change scores for the two measures.

No significant tendency of error reduction was found. In the 49 units of discussion, error (expressed as the sum of the members' absolute errors on an item) was reduced only 22 times. Moreover, no tendency was found for error reduction to be related to the number of people who spoke during the discussion unit.

Although the error is not reduced, the discussion might nevertheless tend to produce a convergence of estimates toward a common level. A tendency was in fact found for the variance among the individual's estimates to be reduced by discussion ( $p = .08$ ).

*Interpretation.* That accuracy was not increased by the discussion is understandable in light of the high degree of similarity observed in the changes of the estimated and actual consensus. The similarity of the changes, in turn, is in part accountable for by the previously observed relationship between the Ss' estimates and their own opinions. Given this relationship, changes in the Ss' opinions should be accompanied by changes in their estimates.

Whatever the reasons may be, it is apparent that not only do the expressions of opinion influence the estimates; they influence the group consensus itself, and in substantially the same manner. The types of conditions upon which this dual effect depends needs to



be explored. The findings of the present study suggest that rather complex interactions may be involved among expressions of opinion, group consensus, and judgments of consensus, and that the common conception of group discussion as a means of increasing the understanding of the group members' views needs qualification.

#### SUMMARY

Six groups of high school students held discussions during which the individual's opinions and estimates of the group consensus were privately taken on the issues considered.

The degree of estimation accuracy achieved was not noticeably different from the accuracy found by other investigators under conditions of no discussion. Tendencies among Ss to overestimate the number of group members who share their opinion similar to ones previously reported, were also found. Even minorities comprised of only about 10% of the group tended to estimate that more than half the group shared their opinion.

The views expressed by the Ss during discussion influenced both the median estimate of consensus and the consensus itself. The

influence on both was highly similar in that (a) the amount of change in the actual consensus was correlated with the amount of change in the estimated consensus, (b) both the estimated and actual consensus tended to change in the direction of the views expressed, and (c) the amount of change in the estimated consensus was not significantly greater than the amount of change in the actual consensus.

Over the course of discussion, the estimates did not tend to become more accurate. There was a slight tendency toward reduction in the variance of estimates among the Ss.

#### REFERENCES

1. CAMPBELL, D. T. *A study of leadership among submarine officers*. Columbus: Ohio State Univer. Res. Found., 1953.
2. CROW, W. J. A methodological study of social perceptiveness. Unpublished doctor's dissertation, Univer. of Colorado, 1954.
3. NEWCOMB, T. *Personality and social change*. New York: Dryden Press, 1943.
4. TRAVERS, R. M. W. A study in judging the opinions of groups. *Arch. of Psychol.*, 1941, No. 266.
5. WALLEN, R. Individuals' estimates of group opinion. *J. soc. Psychol.*, 1943, 17, 269-274.

Received August 16, 1956.

# COGNITIVE PROCESSES AND INTERPERSONAL PREDICTIONS<sup>1</sup>

HOWARD LEVENTHAL<sup>2</sup>

University of North Carolina

**T**HE predictions of an observer about the behavior of another person are a product of the observer's hypotheses or conceptualization regarding the particular individual. The accuracy of his predictions depends upon the validity of the conceptualization. His hypotheses are, in turn, a function of the information available about the person to be judged, and of the judge's typical way of categorizing information or forming concepts about others. These assumptions were developed into three hypotheses which the present study was designed to test.

A more accurate impression or conceptualization of another person should be possible when more information is given the observer about the individual, and a more accurate prediction of his behavior should therefore result. *Hypothesis 1. The more information provided judges about a person, the more accurate are their predictions of his behavior.*

The judge's typical concepts or ways of grouping others determine the kinds of discriminations and consequent predictions he makes about them. Judges with a larger number of concepts for grouping others (hereafter called Complex Judges) should develop more complete and more accurate impressions of an individual than those who use few concepts (hereafter called Simple Judges). *Hypothesis 2. Complex Judges more accurately predict behavior than Simple Judges.*

Since one function of information is to provide the setting for fine or multifactor discriminations, one would predict that the impressions developed by Complex Judges would be more complete and representative of a particular person when they are given additional information. Simple Judges, however, should continue to classify data about the

person observed into the few categories available to them and their impressions should not increase in accuracy with amount of information provided. *Hypothesis 3. With increasing information, Complex Judges show a relatively greater increase in predictive accuracy than Simple Judges.*

## METHOD

The design required a pool of subjects to serve as judges and interviewees, who varied in respect to simplicity or complexity of interpersonal concepts, and in the amount of information received. Judges were given information about another person by means of tape-recorded interviews, and then they predicted the other person's responses to a questionnaire.

### Selection of Judges and Interviewees

Bieri's modification of the Kelly Role Repertory Test (1, 3) was used to rate subjects for conceptual simplicity or complexity. Subjects (Ss) were given a list of 12 different categories of people (e.g., a person you'd like to know better; the most ethical person you know, etc.). The S named an acquaintance or relative to whom each description could be applied. He compared selected triads of these 12 individuals, and for each triad he was instructed to provide a personal characteristic or trait in regard to which two of the persons were similar to one another but different from the third. Any trait or psychological concept could be used for the differentiation. The S then checked the two like individuals. His next task was to see if the concept applied to any of the other nine individuals not in the original comparison, and to check any to whom the concept applied. The Ss made twelve comparisons, using a different specified triad for each.

Scorable protocols on this test were obtained from 253 male students in elementary psychology, and scored for overlap or simplicity in use of concepts. To the extent that the subject checked the same individuals as examples of the constructs he had formed in the successive comparisons of triads, he obtained a high score on simplicity. Scores obtained ranged from 0 to 82, with a distribution skewed toward high, or simple scores. Subjects scoring 20 or above were classified as simple, and those scoring between two and six as complex. From these extreme groups, 14 Ss were selected as interviewees (7 simple and 7 complex), and 56 as judges (28 simple and 28 complex).

To determine whether simplicity-complexity was related to intelligence, the A.C.E. scores for a randomly selected sample of these Ss, 95 in all, were correlated

<sup>1</sup> This study is a condensed version of a dissertation submitted to the University of North Carolina in partial fulfillment of the requirements for the Ph.D. degree in June of 1956. Thanks are extended to Dr. June E. Chance of the Department of Psychology, and to Dr. David Cox of the Department of Bio-Statistics for their invaluable help and criticisms.

<sup>2</sup> Now with the Behavioral Studies Section, U.S.P.H.S.



with their simplicity-complexity scores. The obtained correlation ( $r = -.04$ ) was not significantly different from zero.

### Information Provided

The information from which judges were to make their predictions was obtained by tape recording interviews with seven simple and seven complex individuals. Four content areas were covered in the interview: (a) family background—description of parents and siblings; (b) school and job experiences—descriptions of teachers and bosses as well as relevant activities; (c) values—description of girl friend, and ideal mate, future plans; (d) self-description—how others see self, desires for personal change, and attitudes toward others. The information from topics *a* and *b* above, about 15 minutes of the tape, constituted partial information, while the entire interview of about 35 minutes constituted the condition of complete information.

### The Prediction Questionnaire

Upon the close of the interview, the interviewee gave his answers to a 38-item questionnaire which then served as the criterion for the accuracy of the judges' predictions.

This questionnaire consisted of 38 four-choice items concerning behavior in social, work, and school situations. The questionnaire was administered in the elementary psychology classes to determine the distribution of responses to the items under conditions of self-description. Completed questionnaires were obtained from 236 males in addition to the 56 who later served as judges.

The distribution of responses for the group of 236 showed definite stereotypy of self-description, indicating that the chance level of accuracy would not be the same for all item alternatives. In order to compensate for the advantage this stereotypy would otherwise have given a judge who relied on group norms rather than on the interview information for his predictions, a weighting formula was developed which gave greater weight to accurate predictions the more the interviewee's response deviated from the group stereotype and the more the judge's prediction deviated from the use of "normative concepts." The proportion of times judges predicted each alternative was tabulated for every item, and weights to be applied to correct predictions of an alternative were derived according to the following formula: one minus the proportion of self choices divided by the proportion of times the choice is predicted by judges. As the proportion of times an alternative is used for self-description or prediction increases, the weight decreases. The use of the formula was intended to give greater weight to predictions that reflected discrimination of qualities characteristic of the individual interviewee.

### Design and Procedure

The experiment was designed so that both Simple and Complex Judges judged a simple and a complex interviewee, with partial and complete information, and with each of these combinations of variables in first and second position. The design (Table 1) called

TABLE 1  
EXPERIMENTAL DESIGN SHOWING TYPE OF  
INTERVIEWEE, AMOUNT OF INFORMATION,  
AND SEQUENCE OF JUDGING FOR SIMPLE  
AND COMPLEX JUDGE PAIRS

	Judge Pairs			
	1	2	3	4
Order Judged	Simple Complex	Simple Complex	Simple Complex	Simple Complex
First				
Amt. of Inf.	15 min.	35 min.	15 min.	35 min.
Interviewee	simple	simple	complex	complex
Second				
Amt. of Inf.	35 min.	15 min.	35 min.	15 min.
Interviewee	complex	complex	simple	simple

for four pairs of judges, each pair composed of a Simple and Complex Judge, and one pair of interviewees, a Simple and a Complex. This design was replicated 7 times with a total of 56 judges and 14 interviewees participating.

Two judges, one Simple and one Complex, were scheduled for each of the judging sessions. Before listening to the recorded interviews, judges studied the content of the questionnaire items. They were then instructed to listen to each of two interviews, and to formulate as accurate an impression of the interviewees as possible. After hearing each recording, the judges answered the questionnaire in the way they thought the interviewee had. An incentive award of two dollars was offered to the judge who was most accurate for one interviewee, and an eight dollar prize if he was most accurate for both.

### RESULTS AND DISCUSSION

*Hypothesis 1.* Predictions made with complete information were more accurate than those made with partial information ( $p < .05$ ), when unweighted scores are considered. This effect was not significant for the weighted accuracy scores. Perhaps complete information provided the judges with a clearer idea of the interviewee's subgroup membership, rather than of his idiosyncratic characteristics.

*Hypothesis 2.* There was no difference between Simple and Complex Judges in the total number of correct predictions. However, when compared on weighted accuracy, Complex Judges showed a tendency to be more accurate ( $p < .10$ ). This trend is consistent with the interpretation that Complex Judges are more likely to base their predictions on the characteristics of a particular interviewee.

*Hypothesis 3.* A comparison of the changes in unweighted accuracy scores revealed that Simple Judges showed relatively greater improvement than Complex Judges when

TABLE 2  
ANALYSIS OF VARIANCE, UNWEIGHTED  
ACCURACY SCORES

	df	Variance	F
Between Judge Source			
Between Subject Pairs	6	52.31	3.76***
Judge Type	1	.08	
Information $\times$ Order	1	2.01	
Information $\times$ Subject Type	1	6.51	
Order $\times$ Subject Type	1	16.51	
J. Type $\times$ Subj. Pairs	6	13.00	
J. Type $\times$ Inf. $\times$ Order	1	8.58	
J. Type $\times$ Inf. $\times$ S. Type	1	13.58	
J. Type $\times$ Order $\times$ S. Type	1	1.08	
Residual Between	36	13.91	
Within Judge Source			
Subject Type	1	40.08	
Subj. Type $\times$ Subj. Pairs	6	31.87	4.60***
Information	1	33.22	
Order	1	25.08	
J. Type $\times$ Inf.	1	19.72	
J. Type $\times$ Order	1	33.22	4.92**
J. Type $\times$ S. Type	1	.72	
Inf. $\times$ Order $\times$ S. Type	1	8.58	
Residual Within	43	6.92	
Residual Bet./Residual Within	36/43		2.01**

\*.10 >  $P$  > .05.

\*\* .05 >  $P$  > .01.

\*\*\* .01 >  $P$  > .001.

judging with complete as compared with partial information ( $p < .10$ ). The increase in unweighted accuracy with complete information was, therefore, largely due to increased accuracy by Simple Judges.

These data raise two questions: (a) How did Simple and Complex Judges differ in the process by which they arrived at their predictions, and (b) What role did perception of individual interviewee characteristics play in prediction? How can we understand the trends for Complex Judges to predict more atypical items and for Simple Judges to improve more with complete information? Could these differences in accuracy have been a function of actual similarity between the judges and the interviewees in their own responses to the questionnaire? Analysis of similarity in self-response indicated no significant differences in judge-interviewee similarity for any experimental condition.

Since Bieri (1) found greater "assimilative projection," i.e., prediction of similarity, in Simple Judges, it is possible that prediction of similarity can account for the accuracy of

Simple Judges in the present experiment. An analysis showed significantly greater prediction of similarity between interviewee's and judge's responses by Simple Judges than by Complex Judges ( $p < .05$ ). Therefore, the components of accuracy (the proportion of actual response similarities accurately predicted, and the proportion of actual response differences accurately predicted) were analyzed. The trends showed that Simple Judges correctly predicted a greater proportion of similarities and the Complex Judges correctly predicted a greater proportion of response differences. However, neither trend was significant.

Though Simple Judges predicted significantly greater similarity than did Complex Judges, Complex Judges were more able to order the interviewees in terms of similarity to themselves. Correlations of actual similarity with amount of similarity predicted by the judge were computed separately for each combination of experimental conditions and scores averaged. The correlations were,  $r = .49$

TABLE 3  
ANALYSIS OF VARIANCE, WEIGHTED  
ACCURACY SCORES

	df	Variance	F
Between Judge Source			
Between Subject Pairs	6	260.34	2.66**
Judge Type	1	375.22	
Information $\times$ Order	1	82.05	
Inf. $\times$ Subj. Type	1	12.04	
Order $\times$ Subj. Type	1	30.93	
J. Type $\times$ Subj. Pairs	6	122.24	1.25
J. Type $\times$ Inf. $\times$ Order	1	85.16	
J. Type $\times$ Inf. $\times$ S. Type	1	4.58	
J. Type $\times$ Order $\times$ S. Type	1	40.87	
Residual Between	36	97.77	
Within Judge Source			
Subject Type	1	175.20	
Subj. Type $\times$ Subj. Pairs	6	406.87	4.62***
Information	1	188.45	
Order	1	385.32	4.38**
J. Type $\times$ Information	1	18.14	
J. Type $\times$ Order	1	438.06	4.97**
J. Type $\times$ Subj. Type	1	7.22	
Inf. $\times$ Order $\times$ Subj. Type	1	6.75	
Residual Within	43	88.06	
Residual Bet./Residual Within	36/43		1.11

\*.10 >  $P$  > .05.

\*\* .05 >  $P$  > .01.

\*\*\* .01 >  $P$  > .001.



( $p < .01$ ) for Complex Judges, and  $r = .24$  (not significantly different from zero) for Simple Judges. These results support the possibility that Complex Judges predicted on the basis of an impression of the individual interviewee that was generated from the information provided him. Although this interpretation is in agreement with the finding that Complex Judges were more accurate on individually relevant items, the data are more suggestive than conclusive.

Observing judge performance as a function of the other independent variables may further clarify the nature of the prediction process as well as differences between the Judge types. While the increase in accuracy with full information was greater for Simple Judges than for Complex, both types of judges showed greater accuracy under this condition. The increase in accuracy cannot be attributed to predictions of similarity since predictions of similarity decreased significantly ( $p < .05$ ) with complete information. Complete information also resulted in a decrease in the proportion of real differences incorrectly predicted as similarities ( $p < .01$ ), and an increase in the proportion of correctly predicted differences ( $p < .001$ ). These results may be a function of the content of the information as well as of the amount of information. Complete information alone disclosed the interviewees' self-perceptions to the judges. A possible reason for complete information being more helpful to Simple and Complex Judges is that the material on the interviewees' self-perceptions may have been very relevant for correcting the kind of errors the Simple Judges were making, i.e., not differentiating the interviewees from themselves. On the other hand, Complex Judges, when making predictions with complete information, may have over-differentiated between themselves and others. Over-differentiating the interviewees is likely to result in error, since people tend to respond similarly to the questionnaire. Trends in the data showed that with complete information, Complex Judges decreased their accurate predictions of similarity, and increased their inaccurate predictions of differences. Perhaps Complex Judges would have been more accurate if they knew more about the normative responses to the items, or if the criterion had recorded more individualistic responses.

TABLE 4  
ANALYSIS OF VARIANCE, PREDICTED SIMILARITY  
BETWEEN JUDGE AND INTERVIEWEE

	df	Variance	F
Between Judge Source			
Between Subject Pairs	6	38.85	
Judge Type	1	150.89	4.14**
Information $\times$ Order	1	2.89	
Inf. $\times$ Subj. Type	1	48.89	
Order $\times$ Subj. Type	1	11.57	
J. Type $\times$ Subj. Pairs	6	28.27	
J. Type $\times$ Inf. $\times$ Order	1	14.29	
J. Type $\times$ Inf. $\times$ Subj. Type	1	137.29	3.77*
J. Type $\times$ Order $\times$ Subj. Type	1	22.32	
Residual Between	36	36.44	
Within Judge Source			
Subject Type	1	6.04	
Subj. Type $\times$ Subj. Pairs	6	23.08	1.43
Information	1	89.29	5.53**
Order	1	54.32	3.36*
J. Type $\times$ Inf.	1	26.03	
J. Type $\times$ Order	1	24.14	
J. Type $\times$ Subj. Type	1	20.57	
Inf. $\times$ Order $\times$ Subj. Type	1	36.57	2.26
Residual Within	43	16.15	
Residual Bet./Residual Within	36/43		2.26***

\* .10  $> P > .05$ .

\*\* .05  $> P > .01$ .

\*\*\* .01  $> P > .001$ .

Judging interviews in succession produced effects upon accuracy and predicted similarity. The trends for the second interview were toward decreased accuracy, ( $p < .10$  for unweighted accuracy and  $p < .05$  for weighted accuracy) and greater prediction of similarity ( $p < .10$ ). The decrement in accuracy was largely confined to Complex Judges ( $p < .05$  for the Judge Type  $\times$  Order interaction on both accuracy measures). Simple Judges did about equally well with interviews judged in either position. The deterioration of prediction among Complex Judges may well reflect their inability to continue to respond to the individual interviewee's characteristics. If a Complex Judge tried in his predictions to distinguish the second interviewee from both himself and the first interviewee, the alternatives remaining for his choice were likely to be those least frequently selected for self-description.

In general, it appeared that Complex Judges responded in terms of differences between people, while Simple Judges responded in terms of similarities. Viewing people as similar to oneself as an initial hypothesis may help

Simple Judges to understand and adjust to their own interpersonal environments. They differentiate only when sufficient information is available. Complex Judges, on the other hand, seem to seek a unique characterization of a person and to be fairly able to pick out aspects of his individuality even with little information. It appears that the general tenor of the experimental instructions, which implied judging the uniqueness or differences between individuals, had a stronger effect upon the Complex Judges. The Simple Judges of the present study resemble Lundy's (5) Ss who were instructed to attend to their own reactions while evaluating another person, and Fiedler's (2) Ss who rated high in assumed similarity between opposites.

Do we have data that can tell us whether the judges actually perceived and responded in terms of characteristics of the individual interviewees, as well as responding in terms of their expectations of how college males would answer a questionnaire? The changes in accuracy and predicted similarity with information give support to the view that information as well as expectations determined the predictions. Had data on the effects of information and order not been obtained, our conclusions might have been different. We would have seen that accuracy correlated with similarity, and that variables such as the proportion of similarities and proportion of differences correctly predicted did not correlate, and might have concluded that information specific to an individual interviewee was not significant for accurate prediction.

The results suggest that research directed toward explaining the prediction process should include (a) more specific treatment of judges' self-constructs or the ways in which they compare themselves to others, (b) manipulation of the similarity between judge and judged as well as control of the similarity of their responses to group norms, (c) more careful control of variations in the content of information given to judges, and (d) systematic variation of the relevance of the content of the information to the content of the predictions.

#### SUMMARY

A set of interrelated variables thought to influence predictive accuracy were studied

experimentally. A model treating social perception as an example of concept formation suggested the combined manipulation of stimulus and individual difference variables. It was expected that judges possessing a number of differentiable concepts regarding others (Complex Judges) would be better predictors than judges having undifferentiated interpersonal constructs (Simple Judges). It was hypothesized that given more information about a person, judges would predict better and that Complex Judges would best utilize full information. Scores on the Kelly Role Repertory test defined the dimension of conceptual simplicity-complexity. Information was presented by means of recorded interviews, each type of judge judging each type of interviewee, and judgments were made on the basis of two different amounts of information. The results indicated that though Simple and Complex Judges were about equal in predictive accuracy, they differed in the ways in which this accuracy was achieved. Simple Judges tended to assume similarity between themselves and others, while Complex Judges tended to differentiate between themselves and others. The increase in accuracy resulting from the provision of more complete information also seemed to indicate that the judges responded in terms of the individual characteristics of the interviewees, as well as in terms of group norms.

#### REFERENCES

1. BIERI, J. Cognitive complexity-simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, 51, 263-268.
2. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. In A. P. Hare, E. F. Borgatta, and R. F. Bales (Eds.), *Small groups*. New York: Alfred A. Knopf, 1955, Pp. 252-265.
3. KELLY, G. A. *The psychology of personal constructs*. Vol. I. *A Theory of Personality*. New York: W. W. Norton, 1955.
4. LEVENTHAL, H. Effects of setting factors and judge's interpersonal constructs on prediction behavior. Unpublished doctor's dissertation, Univer. of North Carolina, 1956.
5. LUNDY, R. M. Assimilative projection and accuracy of prediction in interpersonal perceptions. *J. abnorm. soc. Psychol.*, 1956, 52, 33-38.

Received September 19, 1956.



# CONFORMING BEHAVIOR UNDER TWO CONDITIONS OF CONFLICT IN THE COGNITIVE FIELD<sup>1</sup>

FRANZ SAMELSON<sup>2</sup>

*Mental Health Research Institute, University of Michigan*

ONE of the oldest topics in experimental social psychology is the effect of a communication from other persons on an individual's opinion or judgment: suggestibility and conformity. While the early experiments simply demonstrated this effect in various situations, later ones have shown that the amount of conformity obtained is related to three clusters of factors: (a) the task stimulus, its difficulty and ambiguity (5, 14); (b) the social influence, its source, content, and discrepancy with either the true facts or the individual's initial judgment (2, 9); and (c) characteristics of the individual, permanent or temporarily induced (6, 12). While the phenomenon is well enough established—in fact, seems to be one of the cornerstones of social psychology—the interpretation of the change processes, their “rationality” or “irrationality,” has been the subject of controversy (2, ch. 14).

In an attempt to clarify the issue, the distinction between two types of forces, viz., “normative” and “informational” social influences, has been suggested recently by Deutsch and Gerard (7), thus linking the topic with the theory of reference groups (11). The importance of *normative* social influences for conforming behavior is shown in the experiment by Deutsch and Gerard, and a similar one by Mouton *et al.* (16), in which anonymity of the subjects (Ss) reduced the amount of conformity obtained significantly, but by no means completely.

*Informational* social influences, or the creation of a frame of reference, as in Sherif's classical study with the autokinetic phenomenon (21), seemed to be especially important in the realm of “social reality,” i.e., in regard to

issues which are not subject to direct observation or independent test. Yet Asch's famous experiment on conformity (2, ch. 16) and its subsequent modifications (*v.s.*) seem to indicate that even in simple perceptual tasks (including immediate memory) wrong information, contained in other Ss' responses, had some power to change the judgments of the naive observer. The implication seems to be that even simple perceptions may need some “consensual validation.” This point touches on the old philosophical problem of “reality” and knowledge about reality, and the suggestion given is quite similar to the philosophical position<sup>3</sup> that sense experience is transformed into experience of reality “out there” of material objects only after its corroboration by other observers, or after repeated independent observations by the individual, especially in the form of what Allport (1) has called denotational procedures. (Typically, this latter course of action is made impossible for the Ss in conformity and suggestion experiments—a restriction which may seriously limit the generality of the results.) In other words, viewing perceptions as hypotheses (4), one may expect that even after their initial formulation they may still be susceptible to confirmation or infirmation from reports of other observers.

If these premises are accepted and the discussion restricted to the informational component of the influence attempt, the following analysis of Asch's experiment seems plausible. The task stimuli (lines to be compared) were by themselves quite free from ambiguity. Yet this clearness of the perception combined with the similarly unequivocal but contradictory evaluation by the other observers to create an unstable total situation for the naive S. He was, in fact, required to choose between two mutually exclusive messages, yet had no independent basis to discount

<sup>1</sup> This paper is based on a doctoral dissertation submitted to the Psychology Department of the University of Michigan. The writer wishes to express his gratitude to Daniel Katz, chairman of the doctoral committee, and its members, especially Theodore M. Newcomb and Russell A. Clark. The study was carried out under a research training fellowship granted by the Social Science Research Council.

<sup>2</sup> Now at Kansas State College.

<sup>3</sup> For a recent discussion, see the article by Fritz (8) and its distinction between S-statements (statements about sense-data) and M-statements (about material objects).

either of them. Since at the outset the task seemed quite simple, the initial expectation of observer agreement was presumably very high. The surprising discrepancy between the *S*'s perceptions and those of a number of other *S*s, unanimous among themselves, created a conflict in the cognitive field which could not be resolved in a satisfactory manner by simply denying or disregarding the socially transmitted information. Instead, *S*s tried to account for the disagreement by means of additional hypotheses which, because of their *ad hoc* nature, left the *S*s rather disturbed and dissatisfied whether they had yielded or remained independent.

On the basis of these considerations, the following hypothesis was proposed: The amount of conformity in situations of the type studied by Asch is reduced if *S*s are given secondary information which, if utilized, reconciles the contradiction between their own perception of the task stimuli and the majority judgments, and thus enables them to account for the disagreement. This hypothesis is somewhat similar to the (untested) Hypothesis VI of Deutsch and Gerard: "The more uncertain the individual is about the correctness of the judgment of others, the less likely he is to be susceptible to informational social influence in making his judgment" (7, p. 630).

#### METHOD<sup>4</sup>

In order to test this hypothesis, two conditions were designed under which *S*s found themselves in disagreement with a secretly instructed majority. Their task was to identify simple visual stimuli that were presented tachistoscopically but above threshold duration. This method of presentation, rather than Asch's original procedure, was used since it made the intended manipulations more plausible<sup>5</sup>. In the control, or Full Conflict (F), condition, *S*s were put—as were Asch's *S*s—into the conflict situation without any additional information. In the Reduced Conflict (R) condition<sup>6</sup>, the *S*s were first given the

opportunity to infer that their recognition thresholds were lower than those of the other participants. Subsequently, they found themselves in the same conflict with the majority as the control *S*s.

#### Subjects

Fifty beginning psychology students, all male freshmen and sophomores at the University of Michigan, were used as *S*s. They were assigned at random to either of the two conditions. This prior assignment, together with scheduling difficulties, eventually left the two groups slightly imbalanced ( $N_F = 26$ ,  $N_R = 24$ ).

#### Task

Each of these *S*s met with four other *S*s, who were in fact secretly instructed accomplices of the *E*, for an experiment "on the span of apprehension." Their task was to recognize a four-consonant nonsense syllable presented at .2 sec., and to select it from four alternative syllables presented approximately 1 sec. afterwards, and left on the screen till all *S*s had responded. The three distractors had been constructed in such a way that one letter and/or its position was changed. For example, for the stimulus syllable FDMR, the wrong alternatives were: PDMR, PDRM, FDRM. All five participants announced their choice publicly in a fixed sequence by calling out the number of the alternative they believed to be correct. The naive *S* was always fourth to report. Twenty trials with different syllables were given. On twelve, all four accomplices gave the same wrong answer. On the remaining eight trials, spaced in an irregular order over the experiment, they chose unanimously the correct alternative. The sequence of critical and neutral trials is indicated in Figure 1.

#### Manipulation of Conditions

The experimental treatment, designed to provide *S*s in the R condition with information potentially reducing the conflict between the physical and the social stimuli, consisted of the following procedure. Prior to the test trials, the groups in the Reduced Conflict condition were given simulated practice trials. Each of three slides with a five-digit number was shown to the group repeatedly at decreasing speeds, beginning with .01 sec. All *S*s were asked to report, out of sequence, what they had been able to recognize. The accomplices, however, had been instructed to act as though they were unable to recognize anything at the initial fast speed, and to give correct responses only after the naive *S* had given an accurate report, and after the speed of presentation of the same slide had been decreased twice. Through this procedure, the *S*s in the Reduced Conflict condition presumably gained the impression that their recognition threshold was lower than those of the other

<sup>4</sup> For a more detailed description of procedures and results, see (19).

<sup>5</sup> Similar modifications of Asch's procedures have been used effectively by others, e.g., Olmstead and Blake (18).

<sup>6</sup> These titles of conditions refer to the predicted amounts of vacillation between independent and conforming responses. Continued independent behavior,

as predicted for the Reduced Conflict group, may well arouse considerable anxiety, so that an over-all measure of conflict experienced by the *S*s would not necessarily show differences between conditions corresponding to these titles.



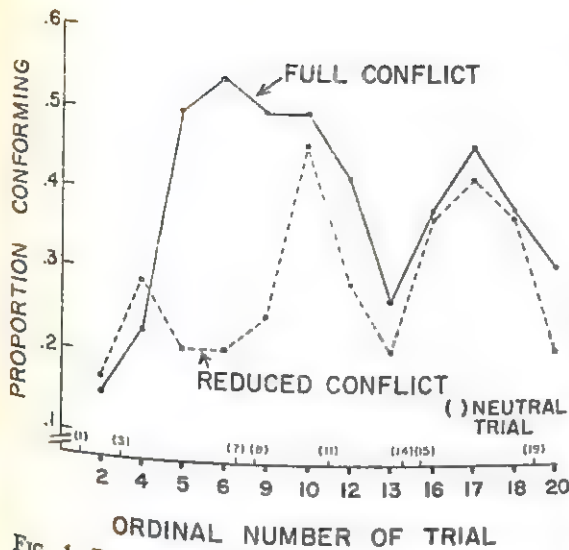


FIG. 1. PROPORTION OF CONFORMING RESPONSES ON SUCCESSIVE CRITICAL TRIALS IN TWO EXPERIMENTAL CONDITIONS

observers. In the Full Conflict condition, no such practice was given. The objective informational as well as normative social influence toward conformity, i.e., the majority's agreement on wrong responses in the test trials, remained identical over both conditions.

#### Apparatus and Setting

A three-hundred watt projector with a Keystone flashmeter attachment was used for presentation of the stimuli. These consisted of four capital letters typed on cellophane and mounted for projection. The distance from screen to projector was 24', to the Ss about 19'. The Ss sat in a row, separated from each other by small screens in order to reduce the demands on the accomplices' acting skill. The back half of the room was illuminated in order to permit Ss to write during the session, as well as to avoid excessive contrast in illumination level.

#### Additional Data

After each judgment, Ss indicated their uncertainty about the correct response on a five-point rating scale from "absolutely certain" to "probably wrong." One of the objectives of this measure was to check, by an intra-individual comparison, whether Ss would differentiate in their ratings between neutral trials and trials on which they had conformed. Since outwardly the response pattern of the group would be identical in both cases, higher uncertainty ratings on the conformity trials would be expected if the Ss had some awareness of conforming to *wrong* majority responses. Such a result would give support to the assumption that Ss did not conform simply because they could not recognize the visual stimuli.

Immediately after the last trial each S was given a questionnaire containing three rating scales and two open-ended questions. The first two ratings dealt with the difficulty of the task and the S's visual acuity. On the third scale he was asked to indicate a judgment of

his ability for a task of the type just performed. This question was designed to test whether the experimental manipulations had been perceived in the intended way. The remaining two items were a question asking whether the S had been influenced by someone else's responses, and a conditional question about his explanation of the disagreement with other Ss "in case such disagreements had occurred."

This task was followed by oral questioning of the naive S under the pretext of a group discussion of the experiment. The topics covered in this interview were the S's awareness of the disagreement with the majority, his feelings about who had been right, and his way of accounting for the discrepancy, gradually leading to the question whether he had developed any suspicions about the behavior of the majority. None of the Ss had been informed beforehand by rumor, and none had surmised the extent of the deception during the experiment, although a few had begun to suspect that the experiment might not be quite what it purported to be. Finally, the purpose and procedures were fully explained to every S since all of them showed some signs of disturbance about the puzzling course of events.

#### Personality Data

Some weeks earlier, in the context of another experiment, measures of the Ss' Achievement and Affiliation motives had been obtained. These projective motivation measures are based on stories written in response to TAT pictures and scored according to conventions developed by McClelland *et al.* for need Achievement (15), and by Atkinson *et al.* for need Affiliation (3). Need Achievement scores are presumably indices of the strength of motivation "concerned with performance in competition with some standard of excellence" (15, p. 111). Need Affiliation, or motivation to be socially accepted, is conceived as "a disposition to move toward others to elicit positive affective responses from them" (3, p. 409).

At another time, the Ss had also filled out the SA-S personality inventory<sup>7</sup>. This questionnaire consists of eighty items drawn from Cattell's Sixteen P F Inventory and presumably measures two orthogonal personality dimensions: anxiety and social approach or "permeability." "Individuals scoring at the permeable extreme tend to be sensitive to stimuli, socially oriented, flexible, and imaginative" (22, p. 324).

Although these personality data had been obtained before the experiment, it had been impossible to evaluate them in time to match Ss in the two conditions. However, subsequent tests showed no significant differences between conditions on any personality measure or intelligence, as measured by scores on the American Council on Education test (ACE).

#### RESULTS AND DISCUSSION

*Differences between conditions.* The first step in testing the hypothesis was to ascertain that the experimental manipulations had been effective. A large difference between conditions

<sup>7</sup> These personality data were graciously made available by Neil A. Carrier and John W. Atkinson.

TABLE 1  
NUMBER OF CONFORMING RESPONSES PER SUBJECT IN TWO EXPERIMENTAL CONDITIONS

Condition	Number of Conforming Responses													N	Mean
	0	1	2	3	4	5	6	7	8	9	10	11	12		
Full Conflict	—	—	9	3	1	5	2	2	—	1	1	2	—	26	4.7
Reduced Conflict	3	6	3	2	1	1	4	2	1	—	—	1	—	24	3.5

on Ss' postexperimental ratings of their ability for the task indicated that the simulated practice had indeed succeeded in structuring the situation differently for the two conditions. Ss in the Reduced Conflict condition, who had had the initial practice, rated themselves significantly more often as "above average" than the Full Conflict Ss ( $p < .001$ ).

Comparing the two conditions on the dependent variable, conformity, we find that the difference is in the predicted direction (see Table 1). However, because of the large variation within conditions, the distributions overlap each other considerably. They are also positively skewed. In order to test the significance of the mean difference, a transformation of the raw data to  $\arcsin \sqrt{\%}$ , as given by Snedecor (23), is appropriate, since the scores represent not true measurements but sums of data in zero-one form (independence of conformity per trial). The resulting  $t$  of 1.81, with 48  $df$ , is significant beyond the .05 level (one-tailed). The null hypothesis of no better performance (i.e., less conformity) in the Reduced Conflict condition can be rejected with reasonable certainty.

Another way of looking at the difference between conditions is given in an analysis by trials. The proportions of conforming responses on each critical trial are shown in Figure 1. On only two trials is the difference in conformity between Full and Reduced Conflict condition negative. A simple sign test shows that the smaller proportion of conforming responses per trial occurred significantly more often in the Reduced than in the Full Conflict condition ( $p < .02$ , one-tailed). A Wilcoxon signed rank test increases the significance of the difference beyond the .01 level.

In addition, Figure 1 shows that on the first two trials only a few Ss in either condition conformed. The direction of the (not significant) difference between the two conditions was even reversed. This result, which was not predicted, helps to reject an otherwise

serious objection to the procedures employed. It indicates that the overall group difference is not due simply to a better adaptation of the Reduced Conflict Ss to the physical features of the task as a consequence of their prior practice with a similar one. Apparently, the effects of the majority response reached their real impact on the Full Conflict Ss only on the third critical trial, while the majority of the Reduced Conflict Ss was still able to withstand them.

The subsequent fluctuations from trial to trial are probably due, in part, to differences between stimuli. These effects are hopelessly confounded in the present design with effects of the spacing of neutral and critical trials<sup>8</sup>. Yet there appears to be a general trend toward decreasing differences between conditions later in the experiment, presumably because the effects of the differential treatment were wearing off. A chi-square test of low vs. high conformity by condition for each half of the experiment results in a significant difference in the predicted direction ( $p = .025$ , one-tailed) in the first half. In the second half, the difference is no longer significant (although by a parametric test it still is). This decrease of the differential effect is not too surprising if we recall that the treatment actually consisted of a rather artificial distortion of the true facts.

A comparison of conditions on the uncertainty ratings indicated a tendency toward lower mean uncertainty for the Reduced Conflict condition ( $.15 > p > .1$ ) failing to reach significance. This lack of difference had been expected, since the scale points were not well enough defined to ensure interindividual comparability. Intra-individual analysis of the different response categories, however, showed

<sup>8</sup> Such effects of the spacing of neutral and critical trials appeared more clearly in the experiment by Asch (2) and one by Olmstead and Blake (18), although in the latter case they were explained as due to differences in the amount of error of the majority judgments.



that most Ss in both conditions had rated their conforming responses on the average as more likely to be wrong than their correct responses in agreement with the majority on the neutral trials. This significant difference ( $p < .001$  with a sign test) supports the assumption underlying the experiment, that Ss would not conform simply because the visual stimuli by themselves were too ambiguous. No consistent differences appeared in a comparison of uncertainties for conforming and independent responses.

The remaining questionnaire responses showed lower ratings for the difficulty of the task in the Reduced Conflict condition, falling short of significance ( $.1 > p > .05$ ), and no differences in the ratings of the Ss' visual acuity and the influence of the majority responses. In answer to this last question, about half the Ss in either condition insisted that they had not been influenced at all by the majority. Their insistence decreased notably when they became aware, during the discussion, that precisely this effect of the majority responses had been the focus of the E's attention.

Beyond this, the interview did not produce any clear-cut results. Although reactions varied considerably, they all indicated that the Ss understandably did not have a clear and consistent picture of the past happenings. Most of them would insist that they had been right whenever they had disagreed with the majority. But in the next sentence they would admit that the judgment of four people had a much higher chance of being correct than their individual one, and therefore they probably had been wrong.

*Personality correlates of conforming responses.* The large variation within both conditions indicated that an analysis of the personality data might be fruitful. On the basis of earlier results (15), need Achievement was expected to be negatively related to conformity. The meaning of the other personality measures available—need Affiliation, social approach, and anxiety—was not specific enough to warrant any predictions.

While the details of this analysis and its tentative interpretation are reported elsewhere (19, 20), the major results will be briefly described here. A correlational analysis showed that, while there was no relation of conformity to need Affiliation and anxiety or to intelligence

(ACE), the other two personality measures interacted with the experimental treatment. In the Full Conflict condition, both need Achievement and social approach were *negatively* correlated with conformity (although not significantly). In the Reduced Conflict condition, however, both measures showed a *positive* relation to conformity, which for need Achievement was significant at the .05 level ( $r = +.45$ ,  $N = 22$ ). The difference between the two conditions with respect to the correlations with conformity was significant at the .05 level for need Achievement, and the .02 level for social approach. The fact that the two measures were obtained from quite different and experimentally independent sets of data (projective stories and questionnaire responses, given at different times) strongly suggests that this interaction effect is not spurious.

An analysis of variance indicated that the interaction can entirely be attributed to the (expected) reduction in conforming responses of the medium and low need Achievement and social approach scores in the Reduced Conflict condition. Contrary to expectations, the high scorers on either measure showed no absolute mean difference between conditions; but their relative position was changed from less to more conformity by the shift of the other subgroup means.

While the over-all outcome of this analysis thus confirmed the prediction, the conclusion about the effects of the experimental treatment has to be modified. The reduction of conflict, achieved by means of an implicit *argumentum ad hominem*, decreased the amount of conformity exhibited by the majority of the Ss. But for the Ss scoring high on need Achievement or social approach, this procedure was not effective. Since there is no indication that these Ss failed to perceive the difference between themselves and the accomplices which the experimental manipulation had attempted to create, the explanation seems to be that the assumed reduction of conflict did not actually occur for these individuals. As they were highly motivated to achieve, their strong motivation to be correct may have kept the conflict between their own perceptions and the majority judgments alive. Thus, they were forced to vacillate between independence and conformity, and thereby to conform more than the other Ss in this condition, who could

make better use of the cues indicating that the majority might be in error.

No entirely satisfactory interpretation of the results obtained with the personality measures has been found<sup>9</sup>. It should be mentioned, however, that the few related experiments on conformity employing contrasting experimental conditions as well as personality measures have shown similar interactions between condition and personality (10, 12, 16).

*Further theoretical considerations.* It will be recalled that the majority of the Ss showed the predicted reduction of conforming responses after they had obtained information enabling them to account for the discrepant majority reports. This information had been implicit in the events of the practice, which employed different stimuli and rather different (shorter) exposure times than the test trials. In this situation, one cannot speak simply of reinforcement of specific responses like the individual norm in the autokinetic situation. Neither was *disagreement* with the majority directly reinforced since in the practice, the accomplices' responses eventually agreed with the S's original one. Of course it can be argued that an internal response of *thinking oneself right* was reinforced (13). However, since generalization of this response along unspecified dimensions must be assumed in order to cover different stimuli and different social situations, the value of this concept seems debatable. The interpretation proposed here is that the strengthening of the Ss' independence involved cognitive processes, especially the Ss' inference about *relative* abilities of the observers, and its application to the new conflict situation in which the Ss found themselves. In other words, cognitive relationships among all three elements of the standard social situation—what Newcomb (17) has called the co-orientation system—were involved.

An exclusive emphasis on cognitive processes, of course, does injustice to the complexity of the total situation. The development of the conflict, and the form that its resolution takes, depends on a number of additional factors. Yet it seems that, at least to some extent, two aspects of the situation are separable. To use the language of communication

theory, it can be said that the receiver, confronted with different signals in different channels, has to make the decision as to which of the two is "noise" and which the correct message. And in order to make this decision he needs additional information. Which message he in turn *transmits* depends on additional considerations, if he realizes that the factually correct message may not be "socially correct," and if he expects feedback from the social environment. In fact, these effects may easily react on the first-stage decision, and in the extreme case the S may give up the attempt to collect independent information and rely entirely on the social channel.

Yet the thesis of this study was that a satisfactory analysis of conformity requires more than the definition of a phenotypic class of responses coinciding with the modal values of the group. Not only, as Asch (2) has argued, must there be a genuine conflict between the different inputs, which is not present in experiments with ambiguous or carefully equated stimuli. The prediction of independent behavior depends on additional information about the source of the conflict. Simply to demand of the S (or to train him) to disregard the socially transmitted information without accounting for it (13) may not be "rational" either, since so much of our knowledge about reality is received through social channels, and "social reality" may not be only a delimitable part of our environment but may, in principle, permeate it throughout.

The large interindividual variation in the present experiment and its relation to motivational variables makes it abundantly clear that a cognitive theory alone is not able to explain the events. The readiness to draw the conclusion that the majority is wrong, as well as the willingness to look for information indicating this possibility, may be, in a wider context, important factors in the determination of conformity or independence.

#### SUMMARY

Fifty Ss were given the task of identifying nonsense syllables presented tachistoscopically above threshold duration, while being confronted with wrong responses of a majority of secretly instructed Ss. In simulated practice trials, half the naive Ss had been given the impression that they had lower recognition

<sup>9</sup> A more extensive treatment of these results can be found in (19, 20).



thresholds than the other observers. This manipulation was intended to reduce the conflict between Ss' perceptions and the majority judgments by implying that the majority might be in error. It was predicted that the amount of conformity in this condition would be less than in the standard condition which provided no information enabling the S to reconcile the apparent contradiction.

The prediction was confirmed. Absence of the difference between conditions on the first two trials indicated that the main result was not due simply to the adaptation of the Reduced Conflict group to the physical features of the task.

The large variation within conditions was found to be related to two personality measures: need Achievement and social approach. The Ss with high scores on these measures performed rather similarly in both conditions, while the predicted difference was obtained through a great reduction of conforming responses by the Ss with medium and low scores.

It was concluded that the creation of a frame of reference permitting the interpretation of the majority responses as in error can reduce the amount of conforming behavior. However, some individuals may not be willing or able to make use of such cues.

#### REFERENCES

1. ALLPORT, F. H. *Theories of perception and the concept of structure*. New York: Wiley, 1955.
2. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
3. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, 49, 405-410.
4. BRUNER, J. S., & POSTMAN, L. Perception, cognition, and behavior. *J. Pers.*, 1949, 18, 14-31.
5. COFFIN, T. E. Some conditions of suggestion and suggestibility. *Psychol. Monogr.*, 1941, 53, No. 4.
6. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, 10, 191-198.
7. DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, 51, 629-636.
8. FRITZ, C. A., JR. Sense-perception and material objects. *Phil. and Phenomenol. Res.*, 1956, 16, 303-316.
9. GOLDBERG, S. L. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, 49, 325-329.
10. HARDY, K. R. The influence of affiliation motivation and social support upon conformity and attitude change. Unpublished doctor's dissertation, Univer. of Michigan, 1955.
11. KELLEY, H. H. Two functions of reference groups. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.) *Readings in social psychology*. (3rd ed.) New York: McGraw-Hill, 1952.
12. KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, 45, 267-285.
13. LASKIN, EVA R. Several methods of preparing individuals to resist social pressure to conform. Unpublished doctor's dissertation, Yale Univer., 1955.
14. LUCHINS, A. S. Social influence on perception of complex drawings. *J. soc. Psychol.*, 1945, 21, 257-273.
15. MCCLELLAND, D. C., ATKINSON, J. W., CLARE, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
16. MOUTON, JANE S., BLAKE, R. R., & OLMSTEAD, J. A. The relationship between frequency of yielding and the disclosure of personal identity. *J. Pers.*, 1956, 24, 339-347.
17. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, 60, 393-404.
18. OLMSTEAD, J. A., & BLAKE, R. R. The use of simulated groups to produce modifications in judgment. *J. Pers.*, 1955, 23, 335-345.
19. SAMELSON, F. Group pressure and incongruity in the cognitive field as determinants of conformity. Unpublished doctor's dissertation, Univer. of Michigan, 1955.
20. SAMELSON, F. The relation of achievement and affiliation motives to conforming behavior in two conditions of conflict with a majority. In J. W. Atkinson (Ed.), *The assessment of human motives*. Princeton: Van Nostrand, in press.
21. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.
22. SMITH, D. E. P., & WOOD, R. L. Verbal satiation and personality. *J. abnorm. soc. Psychol.*, 1956, 52, 323-326.
23. SNEDECOR, G. W. *Statistical methods*. Ames: Iowa State Coll. Press, 1946.

Received October 8, 1956.

# THE REPRESSION RESPONSE TO IMPLIED FAILURE AS A FUNCTION OF THE HYSTERIA-PSYCHASTHENIA INDEX<sup>1</sup>

CHARLES B. TRUAX

*University of Wisconsin*

**T**HIS research was designed in order to test the Freudian theory of repression proper and its relation to scores on the repression index of the Hysteria minus the Psychasthenia scales of the MMPI.

Repression proper (or after-expulsion) is the dissociation or removal from consciousness of anxiety laden material (7). "Removal from consciousness" can be interpreted as simply meaning that the anxiety-laden material is not available for verbalization. Presented in terms of Hullian learning theory (1), the repression hypothesis states that verbal response-producing cues which have been associated with anxiety generate anxiety itself, and that the act of inhibiting the verbal response is learned through the reduction of anxiety. Further, the anxiety should generalize to other verbal responses that are associated with the original anxiety-associated response. Repression, then, should be learned to verbal responses that are associated with the original anxiety-producing response.

If repression is to be regarded as only one of the possible learned responses or mechanisms used to reduce the anxiety drive, however, then not all persons subjected to anxiety may be expected to react by repressing the anxiety-laden material. Erickson (3, 4) has suggested that scores on the Hysteria scale minus scores on the Psychasthenia scale of the MMPI differentiate subjects (*Ss*) that characteristically react to anxiety with the response of repression from those who usually respond in some other manner. Empirically, he found that a high *Hy-Pt* group recalled fewer incompleting tasks and were rated as showing more repression than a low *Hy-Pt* group.

Specifically, then, it is hypothesized that those *Ss* scoring high on this repression index show more repression responses to anxiety than do *Ss* scoring low on it.

In an earlier study by Zeller (15), differential retention for two separate groups following failure-induced anxiety was demonstrated.

<sup>1</sup> The author is indebted to Dr. J. C. Gilchrist for suggestions and criticisms in all phases of the research.

This finding, however, can be satisfactorily explained in terms of a decremental effect of anxiety upon performance in a complex task as proposed by Farber (6).

The test of repression used in the present research involved the determination of whether, once new learning was established, retention for associations involving anxiety-laden material was poorer than for associations involving neutral material *for the same S at the same time*, and whether, when anxiety was removed, there was no differential retention of the same materials.

In a recent study of the effects of anxiety and failure, Truax and Martin (13) found an interaction between degree of *personalization* of failure and manifest anxiety. To avoid this problem in research on failure effects, the present study employed an *implied-failure* technique, involving a minimum of personal interaction between experimenter (*E*) and *S*. This procedure has the additional advantage of reducing the importance of *E* in the failure experience, thus probably lessening the release of *S's* anxiety in the form of hostility toward *E*. In the present study, then, *E* made no comments to indicate *S's* failure. Instead, the anxiety-producing task, which was irrelevant to the experimental task, was simply made so difficult that *S* could not perform well. Failure was thus *implied* by the performance itself, and since the task seemed ego-involving, anxiety could be legitimately inferred.

## METHOD

There were six parts to the present experiment: (a) selection of neutral, emotional, and traumatic words to be used as stimuli in a list of 18 paired associates, (b) original learning of the paired associates, (c) induction of anxiety for the traumatic words, (d) first relearning and recall of all paired associates, (e) removal of induced anxiety for the traumatic words, and (f) second relearning and recall of all paired associates.

The emotional and neutral stimulus words used in the paired associates were selected from a word associa-



tion list of 79 items. The traumatic stimuli were taken from the titles of subtests of an "intelligence test" used to imply failure. The response words used in the paired associates were three-letter nonsense syllables with an association value of 53 per cent taken from Glaze's list (9). Since the response words were neutral nonsense syllables, it was anticipated that conscious suppression of the response would be minimized and thus permit a clearer test of repression.

#### *Selection of Neutral and Emotional Words for Each S*

Words from the word association list were presented on a Gerbrand Tachistoscope for a duration of .5 seconds to each *S* individually. The *S* was asked to respond immediately with the first word coming to mind. Reaction times to the nearest .05 sec. and all verbal responses to each word were recorded. Depressing a telegraph key simultaneously started both the tachistoscope and a standard electric clock; releasing the key stopped the clock at the beginning of *S*'s verbal response.

In addition, Galvanic Skin Response (GSR) changes were obtained during the reaction time interval and were recorded by means of a General Electric photoelectric continuous recorder. An a.c. GSR apparatus designed by Grant (10) with dry silver electrodes attached to the fingers was used. Since the size of GSR changes varies with *S*'s base level of resistance, the sensitivity of the measure for each *S* was based upon the square root of conductance at his base level to permit intra-individual comparisons. The actual sensitivity of the GSR measure ranged between 30 and 600 ohms.

The neutral words for a given *S* were then defined as the six words to which the smallest GSR deflections and short reaction times were given, while the emotional words for a given *S* were defined as the six words to which he gave the largest GSR deflections and short reaction times. Long reaction time words were not used for two reasons: (a) long reaction times is correlated with association or entropy value, and (b) failure to respond immediately results in spuriously large GSRs, possibly because of anxiety over disobeying instructions to respond quickly. As a result of this selection procedure, each *S* had individualized lists of neutral and emotional words, defined by his own responses to them. In addition, the neutral and emotional words for each group of *S*s were equated on the basis of the Thorndike-Lorge word count (13).

#### *Implied Failure and Traumatic Words*

In the session with each *S* between original learning and the first relearning and recall, failure was induced and associated with the traumatic words. Each *S* was given a test booklet with the words "Wisconsin Test of Intellectual Abilities: College Form B" displayed on the upper half of the cover sheet. This test consisted of six subtests prominently labeled with the names to be used as the traumatic words in the paired associate task: *opposites*, *numbers*, *completion*, *reasoning*, *arithmetic*, and *analogies*.<sup>2</sup> Each subtest consisted of ten items of extreme difficulty. At the beginning of each subtest, *E* read the title and gave a brief but glowing ac-

TABLE 1  
EXPERIMENTAL SEQUENCE

Day	Session	Interession Interval	Operation Involved
1	0	—	Selection of emotional and neutral words
2	I	2-3 wks.	Original learning of all 18 paired associates
3	0	24 hrs.	Anxiety induction by implied failure
4	II	24 hrs.	First relearning and anxiety removal
5	III	24 hrs.	Final relearning under neutral conditions

count of the abilities measured by the test, stressing their relationship to general intelligence. The name of the subtest (traumatic word) was repeated three times in the instructions for each subtest. The instructions emphasized speed, and a stop watch was used conspicuously. After each *S* completed only half the items, *E* would click the stop watch and say "Stop." No specific comments about the adequacy or inadequacy of the performance were made.

#### *Disabusal of Failure-induced Anxiety*

At the conclusion of the first relearning session, each *S* was asked to recall the "intelligence test" that he had taken the day before. The *E* asked *S* how he thought he had done on the test and how he felt about it afterward. After listening to *S*'s remarks, *E* continued:

[Actually, that wasn't an intelligence test at all. We are trying to study the effect of failure on verbal learning, so we simply made up a fake test. As you probably noticed, the fake test was so difficult that everyone had to do poorly on it. Usually there aren't any time limits on this sort of test, but to make you feel that you failed we stopped everyone when they were halfway thru. By the way, did you notice that part of the response words that you had to learn were the same as the names of the subtests of the fake I.Q. test? (*E* recorded the responses of *S*) Well, I hope the fake test didn't upset you too much. Tomorrow we'll finally get to the GSR conditioning.]

#### *Learning, Recall, and Relearning*

Each *S* learned and relearned by the method of adjusted learning (8) to a criterion of two successive correct anticipations for each paired associate. The stimulus and the response words were printed on different sides of 3 x 5 cards with the nonsense syllables paired with the stimulus words so that a given nonsense syllable was paired with each of the three classes of stimuli equally often. With this method of counterbalancing, any difficulty due to the nonsense syllables was controlled.

The 18 cards containing the paired associates were shuffled between each trial to eliminate serial position effects. The stimulus and then the response words were presented to *S* separately for a two-sec. period. A one-sec. interval between pairs and a five-sec. interval between successive list presentations was used. After two

<sup>2</sup> Available on request from the author.

successive correct anticipations for a given pair, that pair was discarded, and learning or relearning continued with the remaining pairs.

### Experimental Sequence

The procedural sequence is presented in Table 1. It will be remembered that *S* was led to expect GSR conditioning on the days following the original learning of the paired associates. By this technique, it was hoped that any inter-session rehearsals of the material would be minimized.

### Subjects

Individuals from the upper and lower five per cent on the *Hy-Pt* distribution were selected from 650 students in the beginning psychology course at the University of Wisconsin. These *Ss* were then contacted and asked to serve in a "problem solving" experiment. The first 20 *Ss* from each end of the *Hy-Pt* distribution that were contacted volunteered and were used in the present experiment. Thus, a total of 40 *Ss* served in this study, with seven women and thirteen men in the high *Hy-Pt* group and seven women and thirteen men in the low *Hy-Pt* group. Identical procedures were used for both groups.

### RESULTS

The performance scores used as the dependent variable consisted of difference scores obtained by subtracting the recall or the learning-relearning scores on the six neutral paired associates from the scores for either the six emotional or the six traumatic paired associates. Four sets of difference scores resulted: (a) traumatic recall difference scores, (b) emotional recall difference scores, (c) traumatic learning-relearning difference scores, and (d) emotional learning-relearning difference scores.

Four separate analyses of variance for double classification with repeated measures were performed on the data, corresponding to the four types of difference scores used as the dependent variable (5). In each case, high and low groups based on the upper and lower five per cent of the *Hy-Pt* index were used as the classes, and measures across the three experimental sessions were used as the repeated measures.

The analysis of variance for the recall difference scores and for the emotional learning-relearning difference scores revealed no significant differences. In the case of the analysis for the recall difference scores, the within *Ss* term also proved to be nonsignificant. This lack of individual variation appears to be due to the constricted range of scores

TABLE 2  
ANALYSIS OF VARIANCE FOR RELEARNING  
OF TRAUMATIC WORDS

Source	df	Mean Square	F
High & low groups <i>Hy-Pt</i> index	1	130.34	—
Between <i>Ss</i> within groups	38	321.07	3.70***
Sessions I, II, III	2	651.33	7.52***
High & low groups $\times$ Sessions	2	1,601.74	18.47***
Sessions $\times$ <i>Ss</i> within groups	76	86.72	
Total	119		

\*\*\* Significant beyond the .001 level of confidence.

caused by relearning. The meaningfulness of these null results with recall scores is, therefore, hard to assess.

The results of the analysis of the learning and relearning of the traumatic words in comparison to the neutral words is shown in Table 2. The main effect of learning-relearning sessions is significant beyond the .001 level of confidence, indicating that retention of traumatic words for all *Ss* changed in relation to the neutral words as a function of the operations of anxiety induction by implied failure and the removal of this anxiety by instructions. That is, the traumatic words, which were harder to learn in Session I, became easier to learn by Session III. Further, the interaction of high and low *Hy-Pt* groups by measures across learning-relearning sessions was significant beyond the .001 level of confidence, indicating that the high and low *Hy-Pt* groups responded differently to anxiety induction and anxiety removal.

To evaluate further the interaction of high and low *Hy-Pt* groups by measures across experimental sessions, the Duncan Multiple Range Test was performed on the data (2). Since extreme groups were used in the present study, the .01 level was taken as the acceptable level of significance. The results of this test indicated that the high and low groups did not differ on original learning or on the final relearning of the traumatic words as compared to the neutral words, but did differ in the relearning session after anxiety had been induced (Session II). In Session II, when repression should be manifest, the high group took more trials to relearn the traumatic than the neutral material as compared to either original learning or final relearning. The low



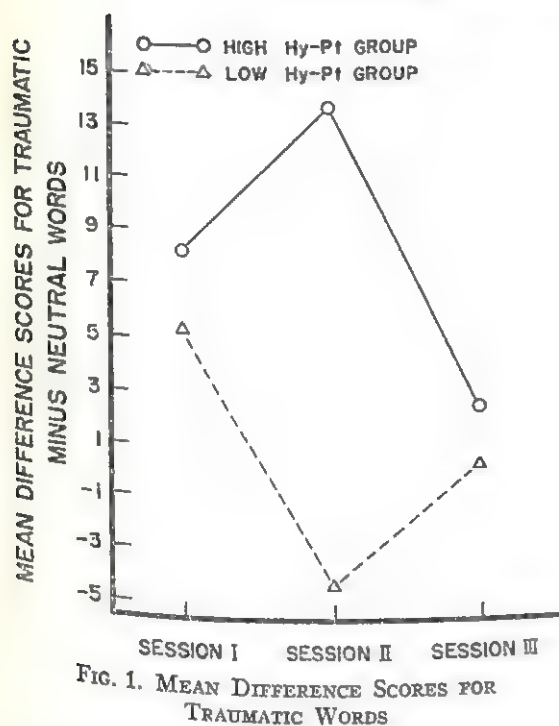


FIG. 1. MEAN DIFFERENCE SCORES FOR TRAUMATIC WORDS

group, on the other hand, took fewer trials to relearn the traumatic than the neutral material in Session II as compared to original learning. The difference between Session II and Session III for the low group, however, fell short of the required significance level ( $p = .06$ ).

In Fig. 1, the mean difference scores for the high and low *Hy-Pt* groups for the traumatic words are plotted as a function of the sessions. Each point represents the trials to the criterion taken by 20 Ss on the six traumatic minus the six neutral words. It will be remembered that Session I was a neutral condition of original learning, Session II the relearning after anxiety induction by the technique of implied failure, and Session III the final relearning after the removal of the anxiety by instructions.

## DISCUSSION

### *The Learning-Relearning Measure of Repression*

When the operation of implied failure is taken as the source of anxiety, the results of the present study are consistent with the Freudian theory of repression proper and the hypothesis that the *Hy-Pt* index differentiates repressors from non-repressors. As shown in Fig. 1, the high group tended to show repressive responses to anxiety, while the low group tended to be facilitated in their performance

by anxiety. That these findings were not the result of suppression, rather than repression, is indicated by the answers to the question: "Did you notice that part of the response words that you had to learn were the same as the names of the subtests of the fake IQ test?" All Ss were quite surprised when this relationship was pointed out to them. If one grants that there was no reason for the S to lie at this point—or even that Ss are not good actors—then the existence of the class of traumatic words was not known by the Ss. It is felt that the 24-hr. interval before and after anxiety induction was crucial to the present study. If the failure had occurred in temporal juxtaposition to the paired-associate task, it seems likely that many Ss would have "seen through" the experiment.

Unfortunately, the results, although showing significant differences in repression, are not free from confounding of the method used with the act of repression. That is, the amount of retroactive inhibition produced by the method of adjusted learning is not known. Since the traumatic words tended to be more difficult in original learning, they were not only presented on more trials, but they were also learned *last*. This, then, allowed for more retroactive inhibition of the neutral words which would tend to make for better relative retention of the traumatic words in Session II. Thus, since repression was not operative in the low group to counteract the effect of retroactive inhibition, these Ss showed faster relearning of the traumatic words in Session II as compared to the neutral words. In Session III, of course, retroactive inhibition would act on the traumatic words since they were relearned first in Session II. If this situation actually did occur, there is no need to postulate an antithetical process such as vigilance to account for the performance of the low group. Whether the results obtained with the traumatic words are due to a process of vigilance in the low group and of repression in the high group or, instead, to a combination of repression and the effects of retroactive inhibition inherent in the method of adjusted learning can only be tested by further research.

When emotional words are taken as the source of anxiety, the lack of significant differences between high and low groups in the retention of anxiety-laden material is

not consistent with past research where differences have been found in relearning scores (14). Such positive results may have been less a function of the effects of anxiety associated with emotional words than of the use of long reaction times as an indicator of the words' emotionality. Emotionality defined in terms of long reaction times tends to be related to association value, which is also measured by reaction time, or to entropy value. Thus, long reaction times indicate low association values and short reaction times high association value for the individual *S*. Studies reporting differences in retention of emotional and neutral words and using reaction time as a measure may have confounded the effects of anxiety and association value. Since no difference in the retention of emotional words was found in the present study where emotional and neutral words were equated in association value by means of (a) reaction time for each *S* and (b) the Thorndike-Lorge word count, this methodological point merits further attention.

#### *The Recall Measure of Repression*

The design of the present experiment did not seem adequate to allow for a clear test of retention of anxiety-associated words in terms of recall scores. A range of only seven scores was possible, and the effect of relearning reduced the observed individual differences to nonsignificance. This aspect of the hypothesis, then, must be regarded as essentially untested.

#### *Implications of the Effect of Implied Failure*

The results indicate that anxiety does develop in a situation where *S* perceives himself as not doing well, even though no knowledge of the adequacy of his performance is given to him by *E*. This suggests that, in studies of learning designed to evaluate differences in the difficulty or complexity of a task, there exists the very real danger of confounding the effects of task complexity with the effects of failure-induced anxiety. It is common for *Ss* to discuss an experiment with others who have already served in it before they volunteer to participate. If the experienced *S* has served in a simpler task and the naive *S* is to serve in a more complex one,

then the experimental situation may imply failure to the naive *S* on the basis of his false performance norms. The resulting confounding of task complexity with the effects of failure-induced anxiety should be particularly prevalent in such studies as those of concept-formation, where the nature of the task remains unchanged but the problem increases in complexity. Thus, for example, the presence of irrelevant information might result in greater time consumption while the occurrence of failure-induced anxiety might result in an inability to complete the task successfully.

To control for the differential presence of implied failure in different levels of task complexity, it is suggested that (a) generous performance norms be given to each *S*, (b) each *S* be told that different *Ss* are given tasks differing in difficulty, and (c) questionnaires be given to *S* at the conclusion of the experiment to evaluate the differential presence of implied failure in groups where task complexity differs.

#### *The Relation Between the Present Results and Manifest Anxiety*

Previous findings, like those reported by Zeller (15), have been susceptible to interpretation in terms of simply the deteriorating effects of failure-induced anxiety. No repression hypothesis seems necessary. In the present study, however, where each *S* acted as his own control and the *Hy-Pl* groups differed significantly in their response to anxiety, a repression process appears to be clearly demonstrated.

The present results cannot be handled by a drive-anxiety interpretation alone. Since scores on the Manifest Anxiety Scale were available for the *Ss* used in the present research, the *Ss* could be divided into high and low anxiety groups. The high *Hy-Pl* group is a low anxiety group (all scores less than 13), while the low *Hy-Pl* group is a high anxiety group (all scores greater than 23). Previous results on the interaction of individual differences in MAS scores and failure anxiety are consistent with the notion that anxiety acts as a drive (6, 13). Considering only the drive aspect of anxiety in the present study, however, would lead to the prediction that the high anxiety group would perform more poorly than the



low anxiety group on Session II (where anxiety is present). The results are in exactly the opposite direction.

While the present results are not at all a contradiction of the drive-anxiety hypothesis, the results do demand that more attention be placed on habits associated with anxiety. It is suggested that the Hysteria variable<sup>3</sup> is a measure of the specific habit of repression as a response to anxiety. Viewed from this perspective, the results are consistent with the clinical notion of repression and with current Hullian learning interpretations of the interaction of drive and habit.

#### SUMMARY

The present research was carried out in order to study the effects of anxiety induced by *implied-failure* in relation to the repression hypothesis and the personality factor involved in the Hysteria scale minus the Psychasthenia scale of the MMPI.

A total of 40 Ss served in the experiment: 20 Ss from the upper five per cent of the *Hy-Pl* distribution and 20 Ss from the lower five per cent. There were six parts to the experimental procedure: (a) selection of six neutral, six emotional, and six traumatic words to act as stimulus words, and 18 medium association value nonsense syllables to act as response words for paired-associate learning, (b) original learning of the 18 paired associates to a criterion of two successive correct anticipations per pair, (c) induction and association of anxiety with the six traumatic words, (d) recall and relearning of the 18 paired associates, (e) removal of the anxiety associated with the traumatic words, and (f) final recall and relearning of the 18 paired associates after removal of anxiety associated with the traumatic words.

The dependent variables used consisted

<sup>3</sup> Little and Fisher (11), in a cluster analysis of the Hysteria Scale found two orthogonal factors; denial of symptoms and admission of symptoms. In rescoring the subscales for the high and low groups used in the present study, it was found that the admission subscale contributed 53% of the Hysteria score for the nonrepressors but only 6% of the Hysteria score for the repressors.

of discrepancy scores obtained by subtracting the trials to reach criterion for the neutral words from trials to criterion for either the emotional or traumatic words. Thus, each S served as his own control.

The results were interpreted as consistent with the repression hypothesis and with the hypothesis that the *Hy-Pl* index is a measure of repressors. The results were also discussed in relation to the drive-anxiety hypothesis and in relation to the implied-failure technique.

#### REFERENCES

1. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
2. DUNCAN, J. Multiple range and multiple F tests. *Biometrics*, 1955, 2, 1-44.
3. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, 47, 230-235.
4. ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, 61, 175-182.
5. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart & Co., 1950.
6. FARBER, I. E., RUSSELL, W., & ANDREAS, B. The effects of failure upon performance in verbal and motor learning situations: A review of work done at the State University of Iowa. Reported to Midwest. Psychol. Ass., April, 1949.
7. FREUD, S. Repression. *Collected Papers*, Vol. IV. London: Hogarth, 1925. Pp. 84-97.
8. GILLETTE, A. L. Learning and retention; a comparison of three experimental procedures. *Arch. Psychol.* 1936, 28, No. 198.
9. GLAZE, A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, 35, 255-267.
10. GRANT, D. A. A convenient alternating current circuit for measuring GSR's. *Amer. J. Psychol.*, 1946, 59, 149-151.
11. LITTLE, K. B., & FISHER, J. The internal structure of the MMPI Hysteria Scale. Paper read at Bay Area VA Clin. Psychol. Res. Meeting, Oakland, Calif., Oct., 1952.
12. THORNDIKE, E. L., & LORGE, I. *A semantic count of English words*. New York: Teachers College, Columbia Univ., 1938.
13. TRUAX, C. B., & MARTIN, B. The immediate and delayed effect of failure as a function of task complexity and personalization of failure instructions. *J. abnorm. soc. Psychol.*, 1957, 55, 16-20.
14. WORCHEL, P. Anxiety and repression. *J. abnorm. soc. Psychol.*, 1955, 50, 201-205.
15. ZELLER, A. F. An experimental analogue of repression: II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, 40, 411-422.

Received October 9, 1956.

# INTRA-INDIVIDUAL CONSISTENCY IN CONCEPTUALIZATION

LOUIS J. MORAN

*University of Texas*

LAURENCE S. MCGAUGHRAN<sup>1</sup> AND DON LEVENTHAL<sup>2</sup>

*University of Houston*

**T**HIS study is concerned with the general problem of whether individuals tend to exhibit a characteristic approach that is consistent and predictable on a variety of conceptual tasks. Both normal and psychiatric groups are used, not for purposes of intergroup comparison, but to examine intra-individual consistency in the conceptual behavior of subjects, whether normal, schizophrenic, or brain-damaged.

For the present purpose, a concept may be defined as a principle under which a variety of particulars may be grouped. These principles may be characterized in a number of ways. For example, a concept may vary in its degree of "publicness," i.e., the degree to which the principle covering a grouping is shared and freely communicated by the majority of persons within the same subculture. Thus, a subject who collects several three-sided figures under the concept *triangles* is using a more "public" principle than a subject who collects the same figures under the relatively private concept, *pleasing*. In other words, a concept may vary in the extent to which it permits public prediction of the objects or events which it is intended to categorize.

Concepts may be classified also in terms of the range or order of objects or events which may be grouped under them. Thus, a subject who characterizes a group of cattle as *organisms* is using a higher order of classification than a subject who uses the concept *animals*; the latter, in turn, is using a higher order of classification than a subject who specifies *cows*. In this instance, the concept *organism* may be called highly "open" in relation to the concept *cows* in terms of the range of other objects that potentially might be included under the principle used.

Concepts may, of course, be classified in a number of other ways. The present analysis,

however, involves only the two variables mentioned, "publicness-privateness" and "openness-closedness." This schema delineates four "conceptual areas" which are represented by the quadrants formed through the intersection of the two variables. Thus, concepts may be classified as public-open, public-closed, private-open, or private-closed. It has been shown that by means of this schema, such groups as the brain-damaged, the schizophrenic, and the intellectually bright and dull may be reliably differentiated (1, 2, 3).

The present study is designed to test the hypothesis that individuals evidence a consistent tendency to operate within a specific quadrant in responding to a variety of conceptual tasks. Such a consistent tendency constitutes a conceptual trait, implying a general and predictable behavior potential toward the solution of conceptual problems.

## METHOD

### *Subjects*

The initial sample consisted of 77 subjects (Ss). Of these, 40 were patients in the Houston Veterans Administration hospital, undergoing treatment for or recuperating from nonpsychiatric illnesses which would not affect in any known way their performance on conceptual tasks. A second group of 37 Ss were backward psychiatric patients in the Waco Veterans Administration Hospital, all diagnosed as chronic paranoid schizophrenics.

### *Procedure*

In a study conducted several years ago by the senior author (4), a battery of seven word-meaning tests was administered to these Ss in order to test several hypotheses concerning the schizophrenic's understanding and use of word meanings. Their performance on two of these tests, a 25-item synonym test and a 17-item similarities test, is reanalyzed here. The new scoring procedures, based upon the conceptual-area schema, were derived a priori. Each item was scored on every test before proceeding to the next item. The consistency with which each S used a given conceptual area on these two tests and on an object sorting test, which was administered at the same time, was tested.

*Sorting test.* The sorting task was administered individually in accordance with the directions given by

<sup>1</sup> Employed part-time at the Houston V. A. Hospital where part of this study was conducted.

<sup>2</sup> Now on the staff of the Topeka State Hospital.



Rapaport (5). In brief, the technique involved the sorting of 33 objects familiar in everyday experience. One part consisted of seven tasks requiring *S* to group with some particular object the others that "belong with it." The second part consisted of twelve tasks requiring *S* to identify a basis for the grouping of a number of objects arranged by the experimenter.

Each of the responses to the sorting test items was assigned to one of the four conceptual areas: public-open, public-closed, private-open, private-closed. Publicness-privateness was scored in terms of the extent to which an individual concept permitted potential public prediction of limits. For example, the boundaries for a concept of "redness" are rather clearly denotable; the conceptual boundaries for "attractiveness" would be considerably less so. The variable of "openness-closedness" was scored in terms of the number of common attributes which were used or implied in each conceptual definition. For example, the classification of a group of objects as "all are red" is "open" relative to such a classification as "all are round and red and rubber." This latter concept binds together more of the perceptible attributes of the objects and, hence, limits the number of other objects that might potentially be included in the same domain. An extended treatment of this scoring system is provided by McGaughan and Moran (2). Interjudge agreement, based upon 55 randomly selected items, was 93 per cent.

**Similarities test.** The similarities test was composed of 17 groups of two to four words which may be categorized under a single concept; for example, *boat, train, car*: "all modes of transportation." The instructions were as follows: "Now I am going to read you some groups of words. I would like you to tell me what the words in the group have in common; that is, in what way they are alike."

A response was scored as "public" if *S* grouped the words into an appropriate category such as *door-window*, "openings"; or described a meaningful similarity among the words like "both part of a house." A response was scored as "open" if *S* responded with a concept that was of a higher order of classification than any of the individual stimulus words; for example, *door-window*, "openings." Interjudge agreement, based upon 200 randomly selected items, was 86 per cent.

**Synonyms test.** The synonyms task was given in two parts. The first part consisted of 25 very familiar words which were read one at a time to the *S* under the following instructions: "Tell me all the words you can think of that have the same meaning as this word. For example, words having the same meaning as *little* are: small, tiny, diminutive, and so on. Now, how many words can you think of that have the same meaning as the word *house*, etc.?"

Performance on this part of the test placed the *S* somewhere on the public-private variable. Responses to each of the 25 words were scored as follows:

- Score 2—gave more correct synonyms than incorrect synonyms.
- Score 1—gave equal number of correct synonyms and incorrect synonyms.
- Score 0—gave more incorrect synonyms than correct synonyms.

Categories *a*, *b*, and *c* represent levels of communicability in *S*'s responses. The scoring is relatively in-

dependent of both *S*'s brightness and the number of responses given. Thus, an *S* may give one correct synonym for each stimulus and still achieve the highest possible score on the public-private variable.

The second part of the synonym task places the *S* somewhere on the open-closed variable. The same 25 words were presented. After each word, however, eight choice words were listed. The number of correct synonyms varied from two to five of the eight alternates. The incorrect alternates were associated in some way with the test word but were not synonyms for it, e.g., house-roof. Subjects were given the following instructions: "After each capitalized word in this list are a number of words in small print. Underline each of the words in small print that has the same meaning as the word in capital letters. Do not guess."

In effect, for each word *S* was given eight concepts to "sort" and to place those together that belonged with the stimulus concept on the basis of synonymy. If he grouped most or all of the correct synonyms, he utilized to the maximum the concept of synonymy and was relatively "open." If he grouped a few of the correct synonyms, he was restricting the concept of synonymy and was relatively closed in his sorting. To the extent that nonsynonyms were also grouped, his concept of synonymy was overinclusive and, hence, was relatively open. A measure of open-closed, then, was given by simply totalling the number of concepts underlined.

In short, each *S* performed on three different conceptual problems. The first task involved the formation by *S* of conceptual groupings with concrete objects. The second task required the formation of a concept to serve as the principle for grouping words. The third task served as a more intensive appraisal of *S*'s use of a single principle, that of synonymy. The degree of consistency in approach to these three tasks was estimated by Kendall's coefficient of concordance ( $W_c$ ).

## RESULTS

Individuals in both groups were reasonably consistent in the nature of their solution on the three different conceptual tasks. On the public-private variable the normal group achieved a  $W_c$  of .47 ( $p < .05$ ) and the schizophrenic group a  $W_c$  of .67 ( $p < .01$ ). On the open-closed variable the  $W_c$ 's were .57 ( $p < .01$ ) for the normal group and .63 ( $p < .01$ ) for the schizophrenic group.

The public-private and open-closed variables appear to vary independently. Rank-order correlations between these two variables did not approach significance with either the normal group or the schizophrenic group.

## FOLLOW-UP STUDY

In an independent study of differences in conceptualization between schizophrenic *S*s and brain damaged *S*s, the same measures were administered to 30 chronic paranoid

schizophrenic patients and to 30 patients with a diagnosis of generalized brain damage at the Topeka State Hospital. At the time the protocols were scored, the scorer was not aware of the hypothesis of intra-individual consistency. Coefficients of concordance were determined after analyses of intergroup differences were completed.

On the public-private variable, the schizophrenic group achieved a  $W_c$  of .78 ( $p < .01$ ) and the brain damaged group a  $W_c$  of .67 ( $p < .01$ ). On the open-closed variable, the  $W_c$ 's for the schizophrenic group and the brain damaged group were .48 ( $p < .05$ ) and .63 ( $p < .01$ ), respectively. In both groups, the public-private and the open-closed variables were uncorrelated with each other.

#### DISCUSSION

In previous studies using the conceptual-area schema, inter-individual comparisons have shown wide differences in approach to conceptual problems. The measurement of individual differences has proved serviceable particularly in the discrimination of deviant groups, such as the psychotic (2) and the brain damaged (3). It was the intent of the present study to test for *intra*-individual consistency on a variety of conceptual tasks. The results of this investigation seem to warrant the conclusion that intra-individual consistency in conceptual-area performance is sufficient to merit the term "conceptual trait."

#### SUMMARY

An object sorting task, a similarities task, and a synonyms task were administered

individually to 40 nonpsychiatric medical patients and to 37 chronic paranoid schizophrenic patients. Performance on each task was scored in terms of a "conceptual area" schema, which placed the subject's conceptual solutions in one of four "areas": public-open, public-closed, private-open, private-closed. Results were evaluated by Kendall's coefficient of concordance. Both nonpsychiatric and psychiatric subjects were found to be consistent in their approach to this variety of conceptual tasks. These results were extended to new samples of paranoid schizophrenic and of brain damaged subjects. It was concluded that the conceptual-area schema provides a means of identifying and measuring conceptual traits.

#### REFERENCES

1. MCGAUGHRAN, L. S. Predicting language behavior from object sorting. *J. abnorm. soc. Psychol.*, 1954, **49**, 183-195.
2. MCGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and nonpsychiatric groups. *J. abnorm. soc. Psychol.*, 1956, **52**, 43-50.
3. MCGAUGHRAN, L. S., & MORAN, L. J. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object sorting. *J. abnorm. soc. Psychol.*, 1957, **54**, 44-49.
4. MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, **67**, No. 20 (Whole No. 370).
5. RAPAPORT, D. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.

Received October 8, 1956.



# DISTINGUISHING CHARACTERISTICS OF COLLABORATORS AND RESISTERS AMONG AMERICAN PRISONERS OF WAR

EDGAR H. SCHEIN, WINFRED F. HILL, HAROLD L. WILLIAMS, AND ARDIE LUBIN<sup>1</sup>

*Waller Reed Army Institute of Research*

AMERICAN soldiers captured during the Korean conflict collaborated with the Communists in varying degrees. Most prisoners of war (POWs) committed trivial acts such as signing peace petitions, but a smaller number of men engaged in persistent collaboration which included writing, signing, and soliciting signatures for peace petitions, delivering anti-United Nations lectures to other POWs, aiding in the preparation and distribution of communist propaganda, encouraging fellow POWs to collaborate, informing on other POWs, and generally aiding the Chinese in their indoctrination program (10). It was not clear from the accounts of the repatriates themselves whether one should seek the causes of such collaboration in the pattern of stresses which the men faced, or in personality and background factors in the men themselves, or in some combination of these circumstances. Many hypotheses have been proposed, varying in the degree to which they attribute collaboration to the so-called communist "brainwashing" techniques, or blame it on "weaknesses" in the men themselves, or on weaknesses in American society (3, 5, 6, 7, 8, 10).

The present study is an attempt to determine whether factors in the background and personalities of the repatriates are related to collaboration. However, it should be pointed out that the search for determinants in the background and personalities of the men does not represent a lack of emphasis in our thinking on the role of the stressful environment (as described in references 4, 8, 11). Rather, it is felt that whatever personal determinants emerge will illuminate the complex relationship between specific situational stresses and the individual upon whom they impinge.

## METHOD

### *Variables and Instruments*

It was decided to investigate background factors, intellectual variables, and personality variables. How-

ever, we could not expect to test any man for more than two hours, which resulted in a somewhat arbitrary set of decisions concerning tests and prevented an extensive sampling of personality dimensions.

In order to determine whether such factors as *home background* and *social status* related to collaboration, the following biographical information was obtained from each man: (a) military rank, (b) whether the subject was in the regular army, reserves, National Guard, or had been drafted, (c) location of pre-service home community, (d) age, (e) years of education, (f) civilian occupation, (g) length of army service prior to capture, (h) length of internment, (i) religion, (j) whether the subject had spent most of his life in a large city, small city, small town, or farm, and (k) whether he had spent most of his life with both parents, one parent, or neither parent.

In order to determine whether level of *intelligence* related to collaboration, the information and comprehension subtests from the Wechsler-Bellevue Adult Intelligence Scale and the two halves (vocabulary and abstraction) of the Shipley-Hartford test (years of education may also be included with these intellectual variables) were administered.

The selection of suitable *personality* dimensions presented a more difficult problem. Because collaboration is probably a very complex multidetermined behavior pattern, specific hypotheses were not formulated. However, it seemed likely that a man's relationship to authority figures should play a large role in determining his response in the prison situation. On the one hand, it might be expected that opportunistic disregard for *any* form of authority would produce a pattern of collaboration. On the other hand, a rigid dependency on *any* form of authority might also lead to collaboration if the POW accepted his Chinese captor as the primary source of authority. These two possibilities led us to administer the Psychopathic Deviate Scale of the Minnesota Multiphasic Personality Inventory (MMPI) and an authoritarian scale resembling the California F scale (1). The F scale could not be used because of the lower educational level of the subjects, and because a number of the F scale items involve content which we considered to be too emotion laden for the POWs (our items put primary emphasis on intolerance of ambiguity). Our scale and the F scale were subsequently administered to a sample of 74 army enlisted men (not former POWs) and were found to correlate +.81 with each other.<sup>2</sup>

Army Institute of Research. Dr. Schein is now at the Massachusetts Institute of Technology and Dr. Hill is at Harvard University.

<sup>2</sup> To save printing costs, this authoritarian scale has been deposited with the American Documentation Institute. Order Document No. 5298 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting

<sup>1</sup> At the time of the study, Capt. Schein, Capt. Williams, and Sp. 3 Hill were at the Walter Reed

TABLE 1  
PER CENT OF REPATRIATES IN THE THREE CRITERION GROUPS

Group	Test Sample		Total Population	
	N	Percent	N	Percent
Resisters	39	5.1	138	4.9
Neutrals	608	80.1	2235	78.6
Collaborators	112	14.8	470	16.5
Total	759	100.0	2843	100.0

As an additional general measure the short form of the MMPI was administered, based on a factor analysis of the full MMPI by Welsh (2, 12). This scale permitted us to obtain a general index of anxiety or neuroticism (Welsh's *A* scale) and a general index of repression, social inhibition, and denial (Welsh's *R* scale). The biographical questions, intelligence test items, and personality scale items were mimeographed for group administration.

### Subjects

The instruments just described were administered to a sample of 759 repatriated army POWs (22 officers and 737 enlisted men), or about one-fourth of the total 2843 army POWs repatriated. These men were tested in groups of 10 to 15 on board three ships bound from Korea to the United States. As far as possible, techniques of systematic randomization were used in selecting the ships and selecting the sample of men on each ship. Subjects (*Ss*) were told that the testers were interested only in general information about the ordeal of imprisonment and the kind of people who were able to survive it, not in the responses of particular individuals. Testing conditions were not ideal because of crowded quarters, occasional seasickness, interruptions, and a general state of guardedness in the *Ss*. However, in the main, motivation was considered adequate by the psychologists administering the tests on each of the ships.

### Criterion of Collaboration

The criterion information regarding collaboration was obtained from the army. Only those men were listed as collaborators against whom the army had reliable deleterious information and whose cases warranted some form of disciplinary action. It is quite possible, therefore, that some of the men classified here as noncollaborators or "neutrals" actually had collaborated to some extent. It should be noted that this classification is based on overt collaborative behavior and does not necessarily imply any ideological conversion.

A few of the noncollaborators were classified as "resisters" on the basis of their prison camp behavior. This group included those men whose army dossiers indicated that they were considered by their Chinese

captors as "reactionaries" and were segregated from other POWs, and those men who were recommended by the army for decorations for their behavior in prison camp.

Thus, the final criterion consisted of three categories: (a) the resisters, (b) the neutrals (those who neither resisted nor collaborated), and (c) the collaborators. The distribution of these three criterion groups is shown in Table 1. As can be seen, this distribution closely approximates the distribution of these three groups in the total population of released army POWs. No significant biases appear to have been introduced by the sampling procedures.

### Analysis

An analysis of variance was performed on each of the quantitative variables. From each of these there was obtained both an eta coefficient, indicating the degree of relationship between the given variable and the criterion, and an *F* ratio, from which the significance level was obtained. Whenever this was significant, *t* tests were performed to determine whether any of the criterion groups differed from each other. Because of the unequal numbers of *Ss* in these groups the *t* test procedure was used, in spite of the fact that the probability levels can not be interpreted exactly. However, the fact that we have only three groups to compare reduces the likelihood of a marked bias by the use of this technique.

Chi squares were computed for all the variables which were in categorical form (nonquantitative).

### RESULTS

The results on all the quantitative variables are shown in Table 2. The most striking feature of these results was that the resisters and collaborators always deviated from the neutrals in the same direction. Moreover, in every case in which the relationship was significant, both of these extreme groups were higher on the variable than the neutrals.

It will be observed that the biographical variables are the most discriminating: all three are significant at the .001 level. Intellectual variables are fairly discriminating, and all but education are significant. Only one personality variable is significant, the *Pd* scale of the MMPI, but its relationship to the criterion is higher than that of any other variable except length of internment. Scores on the authoritarian scale fail completely to discriminate the groups.

Although many of these variables are highly significant (frequently far beyond the cutoff for the .001 level), the eta coefficients are uniformly small. The most powerful predictor actually accounts for less than four per cent of the variance of the criterion. Hence, for practical purposes this is essentially a catalogue



of negative results: a number of variables discriminate significantly, but none of them discriminates importantly.

Moreover, between any two criterion groups, even the statistical significance is not striking. Of the eight variables that show overall significance, only five significantly differentiate resisters from neutrals, and six differentiate collaborators from neutrals. The significance levels of these comparisons are shown in the last two columns of Table 2. No comparisons between resisters and neutrals are shown, as none of these was significant.

Among the *categorical* variables, all of which were of the biographical type, only one reached even the .05 level of significance when tested with chi square. This was type of enlistment, in which the same pattern appeared as in the quantitative variables, there being a higher percentage of both resisters and collaborators among the regular army men than among the draftees (officers, reservists and, National Guardsmen were not included in this analysis). Neither army rank, civilian occupation, religion, number of parents present in the home, rural vs. urban background, nor location of home discriminated the criterion significantly. Even the one significant qualitative

TABLE 2  
RELATION OF QUANTITATIVE VARIABLES TO CRITERION

Variable	Criterion Group			Eta	Significance Level		
		R*	N*	C*	F ratio	R vs. N	C vs. N
Length of internment (months)	$\bar{X}$	31.0	26.6	31.9	.19	.001	.001
	s	5.9	11.0	5.9			
Length of service (months)	$\bar{X}$	50.4	30.3	35.5	.15	.001	.05
	s	57.8	28.1	37.0			
Age (years)	$\bar{X}$	27.0	24.1	25.2	.16	.001	.05
	s	9.1	3.9	4.8			
Education (years)	$\bar{X}$	9.5	9.2	9.5	.06	NS	NS
	s	2.5	2.0	2.1			
Information score	$\bar{X}$	14.4	12.3	13.3	.15	.001	.001
	s	3.5	3.7	3.9			
Comprehension score	$\bar{X}$	10.4	9.2	9.7	.09	.05	NS
	s	3.9	3.2	2.8			
Vocabulary score	$\bar{X}$	22.8	20.5	23.1	.13	.001	NS
	s	8.0	7.3	7.6			
Abstractions score	$\bar{X}$	20.5	19.8	23.1	.11	.01	NS
	s	10.3	9.7	10.2			
MMPI Pd scale	$\bar{X}$	18.2	16.7	18.9	.19	.001	.05
	s	4.3	3.9	5.4			
MMPI A scale	$\bar{X}$	11.2	9.8	11.0	.08	NS	NS
	s	7.4	6.5	6.7			
MMPI R scale	$\bar{X}$	16.3	16.9	16.8	.03	NS	NS
	s	5.1	4.7	4.7			
Authoritarian scale	$\bar{X}$	25.1	25.5	25.1	.03	NS	NS
	s	4.1	5.1	4.9			

\* R stands for resisters, N for neutrals, and C for collaborators.

TABLE 3  
RELATION OF QUANTITATIVE VARIABLES TO CRITERION  
AMONG MEN INTERNED 31 MONTHS AND OVER

Variable	Criterion Group			Eta	Significance Level		
		R	N	C	F ratio	R vs. N	C vs. N
Length of service (months)	$\bar{X}$	57.3	33.4	41.7	.19	.001	.05
	s	61.5	26.7	41.1			
Age (years)	$\bar{X}$	28.0	24.4	25.6	.22	.001	.001
	s	5.2	3.9	5.0			
Education (years)	$\bar{X}$	9.4	8.8	9.5	.15	.01	NS
	s	2.4	1.8	2.2			
Information score	$\bar{X}$	14.6	12.0	13.5	.21	.001	.001
	s	3.6	3.5	4.0			
Comprehension score	$\bar{X}$	10.6	9.0	9.8	.15	.01	.05
	s	4.2	3.0	2.9			
Vocabulary score	$\bar{X}$	23.3	19.6	22.9	.20	.001	.02
	s	7.2	7.2	8.2			
Abstractions score	$\bar{X}$	19.1	18.6	22.4	.15	.01	NS
	s	10.7	9.3	10.4			
MMPI Pd scale	$\bar{X}$	18.8	17.0	18.9	.18	.001	.05
	s	4.6	4.0	5.7			
MMPI A scale	$\bar{X}$	11.1	10.2	11.1	.06	NS	NS
	s	7.9	6.5	7.1			
MMPI R scale	$\bar{X}$	16.4	17.0	16.6	.04	NS	NS
	s	5.1	4.6	4.9			

variable, type of enlistment, was probably artifactual, as will appear later.

Because *length of internment*, which is essentially a situational variable, was so high on the list of predictors, it seemed that a purer measure of the effect of the other variables might be obtained by holding it constant. We therefore re-analyzed a number of variables within the subsample of 457 men interned 31 months and over (i.e., those captured in the first half-year of the war). Table 3 shows the results of this re-analysis. For no variable is the trend changed, and for the majority of variables the magnitude of the relationship is somewhat increased. Education, in particular, now becomes statistically significant. Moreover, for the first time, there appears a significant difference between resisters and collaborators (not shown in the table) on one variable, age. The resisters are older, on the average, than the collaborators. All the etas, however, are still quite small and of no practical value. Scores on the authoritarianism scale were not re-analyzed because of its failure to discriminate in the total sample.

Re-analysis of the one significant categorical variable (enlistment type) in the restricted sample proved to be trivial, because only one per cent of the long-interned enlisted men were draftees, making the use of chi square both impossible and superfluous. The apparent

difference between enlistment types on the criterion is thus easily explained by the longer average internment of regular army men than of draftees.

A number of intercorrelations were found among the predictors. The intellectual variables tended to be positively related both to age and length of service (themselves correlated, of course), and to the *Pd* scale of the MMPI. However, a covariance analysis indicated that the *Pd* scale still discriminates almost as well when any of the intellectual variables is held constant. The information score also remains significant when age is held constant. In view of the small predictive power of all the variables, a more complete analysis of their interrelations did not seem worthwhile.

### DISCUSSION

Perhaps the most important finding in this investigation is negative—none of the variables tested predicted the criterion to any useful extent. Thus, our results argue against any simple explanation of resistance or collaboration in terms of easily measured background or personality variables. For *practical* purposes this study might best be summarized by saying that resisters, neutrals, and collaborators do not differ appreciably on any of the variables investigated.

Particularly striking is the failure of our authoritarianism scale to discriminate the groups. Apparently, the kinds of personality dimensions measured by this instrument are not related to prison camp adjustment, unless one assumes that the repatriates were more guarded and inhibited in answering these items than MMPI items. As we stated previously, we deliberately attempted to remove items that might arouse defensiveness, but we have no way of knowing the extent to which this effort was successful.

Similarly, occupation, number of parents, location of home, rural-urban origin, and religion failed to discriminate the groups, though it is possible that the failure to obtain significant differences may be due to the lesser power of the chi-square test.

Nevertheless, a number of statistically significant differences were found, and theoretically these pose interesting problems. As may be recalled, it had been our hunch that col-

laboration might be related to opportunistic disregard of authority. We hoped to measure this factor by means of the MMPI *Pd* scale and obtain the expected relationship. But why should resisters also have come out high on this scale? This finding suggests that we may have to interpret the *Pd* scale somewhat differently. The scale was standardized on psychopathic deviates, who are characterized by several major adjustment tendencies such as difficulty with authority figures, delinquency, opportunism, and under-inhibition of impulses. In the absence of other information it is difficult to rule out any of the above tendencies as explanations for behavior in prison camp. However, it seems reasonable to suppose that since *both resisters and collaborators* were high on *Pd*, it is the under-inhibition of impulses that is common to these two groups.

Those men who were not able to inhibit impulses when under stress may have been forced by their own acting-out tendencies into taking some kind of stand with respect to Chinese indoctrination pressures: they either gave in to them or arbitrarily resisted them.

It will also be recalled that both resisters and collaborators tended to be older, better educated, more intelligent, and more experienced in the army. This combination of factors suggests that "self-confidence" may have been an important element in determining whether a man would take some action: either collaboration or resistance. If this explanation is correct one would also expect the extreme groups to have higher rank. In fact, the resisters and collaborators did tend to have higher rank, but the relationship did not reach statistical significance.

These two factors, acting-out tendencies and self-confidence, suggest that the underlying behavioral continuum to which the various predictors are related is not *resistance-collaboration* but *action-inaction*. Consistently, the chief contrast is between the neutrals, who defended themselves with a wall of apathy (as described psychiatrically in 11), and the two other groups, both of which were characterized by some positive action toward the Communists. This action might take the form of open resistance, of pseudo-collaboration which it was vainly hoped could be used as an undercover means of resistance, or of genuine collaboration regarded as the best way of



dealing with an otherwise hopeless situation. Those men who were low in self-confidence or in acting-out tendencies would be more hesitant to commit themselves either to open resistance or to collaboration, but would tend to follow the passive middle road of the majority.

A recently completed study of former prisoners of war (9) indicates that a higher percentage of resisters and collaborators than of neutrals had been entertainers, were athletically inclined, and were married. If these findings can be interpreted as reflecting tendencies to respond actively rather than passively to situations, they support our interpretation.

The degree to which ideological change was the basis of collaboration is difficult to assess, because none of the tests used here measured such a variable, nor was there evidence concerning a man's political attitudes prior to capture. It should be noted that there is no *necessary* relationship between collaboration and ideological change. It is possible that ideological change preceded and caused collaboration, or that it followed as a consequence of collaboration, or that the two were entirely unrelated, or that both are a function of some other determinant. Furthermore, data from interviews obtained by the senior author (10) suggest that very few, if any, of the men underwent a significant degree of ideological change. Our results are consistent with the interpretation that ideological change was *not* an important determinant of prison camp behavior. Instead, it appears that the tendency of a person to deal with a stressful situation either actively or passively is a more important determinant of his course of action.

#### SUMMARY

A sample of 759 American army prisoners of war was given a series of biographical questions, intelligence scales, and personality inventories shortly after their repatriation from communist prison camps in Korea. Comparisons were then made among those men who had (a) collaborated with the Communists, (b) actively resisted them, and (c) taken a "neutral" course between these two extremes. Resisters and collaborators were both significantly higher than neutrals in

length of internment, length of service, age, intelligence, and the *Pd* scale of the MMPI, but resisters and collaborators did not differ significantly from one another. No significant differences among the groups were found in army rank, civilian occupation, religion, location of home community, or number of parents present in the home. The *A* and *R* scales from the MMPI and a questionnaire designed to measure "authoritarian" attitudes also failed to discriminate. Although differences obtained were too small to be of any practical value, they are of theoretical interest and were interpreted in terms of an activity-passivity hypothesis concerning response to stress.

#### REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. (Rev. ed.) New York: The Psychological Corporation, 1951.
3. HUNTER, E. *Brainwashing: the story of the men who defied it*. New York: Farrar, Straus, & Cudahy, 1956.
4. LIFTON, R. A. Home by ship: reaction patterns of American prisoners of war repatriated from North Korea. *Amer. J. Psychiat.*, 1954, **110**, 732-739.
5. MAYER, W. E. Why did many GI captives cave in? *U. S. News and World Report*, 1956, Feb. 24, 56-72.
6. MEERLOO, J. A. M. *The rape of the mind*. Cleveland: World, 1956.
7. PASLEY, VIRGINIA. *21 stayed*. New York: Farrar, Straus, & Cudahy, 1955.
8. SEGAL, H. A. Initial psychiatric findings of recently repatriated prisoners of war. *Amer. J. Psychiat.*, 1954, **111**, 358-363.
9. SEGAL, J. Factors related to the collaborative and resistive behavior of the U. S. Army POWs in Korea. George Washington Univer.: Human Resources Res. Office. (To be published in Human Resources Res. Office Tech. Rep.)
10. SCHEIN, E. H. The Chinese indoctrination program for prisoners of war. *Psychiatry*, 1956, **19**, 149-172.
11. STRASSMAN, H. D., THALER, MARGARET, & SCHEIN, E. H. A prisoner of war syndrome: apathy as a reaction to severe stress. *Amer. J. Psychiat.*, 1956, **112**, 998-1003.
12. WELSH, G. S. A factor study of the Minnesota Multiphasic Personality Inventory using scales with item over-lap eliminated. *Amer. Psychol.* 1956, **7**, 341-342. (Abstract)

Received October 12, 1956.

# THE RELATIONSHIP OF AGE AND PRIVILEGE STATUS TO REACTION TIME INDICES OF SCHIZOPHRENIC MOTIVATION<sup>1</sup>

GERALD ROSENBAUM, JAMES L. GISELL, AND WILLIAM R. MACKAVEY<sup>2</sup>

*Wayne State University and Northville State Hospital*

A PREVIOUS experiment (9) presented evidence of a schizophrenic deficit in social motivation on a reaction time task. It was found that biological motivation aroused by means of shocks to a reacting finger tended to eliminate the characteristic slowness and variability of schizophrenic performance. Some of the more disorganized schizophrenics, however, were unable to improve significantly even under shock motivation. The presence of this lessened biological reactivity in some of the patients led to the rejection of the postulate of a unitary deficit in the social motivation of schizophrenics and to the consideration of a multiple deficit conception of schizophrenia. Specifically, it was hypothesized that lessened biological reactivity constitutes a dimension of schizophrenic motivation that is relatively independent of the previously reported deficit in social motivation.

The present experiment was designed to explore this hypothesis further by investigating the relationship of age and privilege status to reaction time under shock motivation. If it could be shown that slow reactions to shock are related to age and not to poor socialization, defined as a lack of ground privileges, then defective biological motivation would be partially identified as a relatively independent dimension of schizophrenic deficit. Similarly, the independence of the dimension of defective social motivation would be indicated by a relationship between privilege status and reactions to social stimulation, essentially unaffected by age. Finally, it was hypothesized that young, privileged patients may not exhibit deficit on either of these dimensions, but may be differentiated from normals by displaying defective reactivity to unpredictable stress as a reflection of disabling anxiety motivation.

In order to test these hypotheses, schizophrenic subjects (Ss), subdivided for age and privilege status, and a group of normal Ss were tested on a reaction time task in which they were required to lift a finger from a telegraph key in response to a buzzer. All Ss were tested under shock, anxiety, and social motivation conditions. In addition, the Henmon-Nelson, a group paper-and-pencil intelligence test (4), was administered to all Ss to obtain some indication of their current mental status. Since Shakow (11) and others (10) have suggested that paper-and-pencil intelligence tests indicate cooperativeness and the ability to sustain attention when used with schizophrenics, the Henmon-Nelson scores were employed as an index of general mental efficiency in assessing the construct validity of the reaction time measures of motivation. The reaction time and Henmon-Nelson tests were later readministered to the patients to determine the reliability of these indices within a schizophrenic population. It was hypothesized that the reaction time scores obtained under each of the motivational conditions would be significantly related to age, privilege status, and the normal-patient classification in the manner indicated.

## METHOD

### *Subjects*

Forty female schizophrenic patients from Northville State Hospital were employed as the experimental Ss, since females had shown the least improvement with shock motivation in the previous experiment. Two conditions of age and privilege status were arranged in a factorial design by subdividing 20 patients over 40 years of age and 20 under 40 years into two equal groups, each consisting of 10 patients with ground privileges and 10 without privileges. Ten college students from Wayne State University, five men and five women, served as normal controls. Since young students, particularly males, are likely to perform better than schizophrenic females on a reaction time task (9), this control group was employed primarily to obtain an index of optimal functioning against which the schizophrenic groups could be compared. The mean ages in years for the groups were as follows: unprivileged-old 49.7, range 43-56; privileged-old 51.2, range

<sup>1</sup> This study was supported by a grant from the Research Division, Michigan Department of Mental Health.

<sup>2</sup> Now at Michigan State University.



41-59; unprivileged-young 28.7, range 18-37; privileged-young 28.8, range 23-35; normal controls 20.2, range 17-25.

### Apparatus

The reaction time device previously described (9) was modified so that the stimulus presentation, recording, and response units could be housed in a portable wooden cabinet 24½ in. long, 12 in. wide, and 12 in. high. The response unit consisted of a telegraph key and a copper plate mounted on a folding panel hinged to the S's end of the cabinet. When the response panel was opened flat on a table, a 3 x 3 in. opal flash screen mounted on the front face of the cabinet was visible. The illumination of this screen by a 10-w. bulb inside the cabinet served as the ready signal.

The ready signal was presented by experimenter's (E's) closing a push-button type switch. Depression of the telegraph key by S in response to the ready signal activated a Hunter Decade Interval Timer and closed a holding relay, permitting E to release his switch. At the termination of the interval, previously set on the interval timer by E, a door-bell buzzer and Standard Electric Timer were instantaneously activated. Release of the key in response to the buzzer broke the common circuit to the buzzer and the clock. Reaction times were read directly to the nearest .01 sec. The clock, interval timer, and all switches were mounted on the rear face of the cabinet, and were accessible when a folding panel was opened.

The shocks, generated by a Harvard Inductarium located inside the cabinet and wired in parallel with the buzzer and the clock, were administered through a copper electrode on the surface of the response key and the copper plate on which S rested his hand. The inductarium was adjusted to transmit a 50-v. current, the setting of a silent mercury switch determining whether or not the shock would accompany the buzzer.

### Procedure

All Ss were instructed that they were taking a reaction time test requiring them to depress the key to the light signal and to release it as rapidly as possible at the sound of the buzzer. The test series consisted of 80 successive trials involving social, anxiety, and shock motivation presented in that order.

**Social motivation.** The Ss were required to respond to the buzzer presented at a constant 10-sec. preparatory interval for 15 successive trials, followed by 10 additional trials at this interval during which S was told his last reaction time and asked to improve his score. The mean of the first 15 reaction times served the individual social motivation scores.

**Anxiety motivation.** The Ss were instructed to be prepared for long and short ready intervals as well as occasional shock, and then presented with 20 trials in which the 1- and 10-sec. intervals occurred equally often in a randomized order. Twenty more irregular trials in which shock was presented on three of the ten 1-sec. trials and three of the ten 10-sec. trials in an unpredictable fashion followed immediately. The mean of the seven reaction times on the 1-sec. nonshock trials served as the anxiety motivation score because it was anticipated that these trials would be most

likely to reflect disruptive anxiety in long reaction times.

**Shock motivation.** The Ss were informed that they would feel skin stimulation at the same instant the buzzer sounded on all of the last 15 shock trials presented at the regular 10-sec. interval. The mean of these 15 reaction times served as the shock motivation score, to be compared with the 15 trials of the test series given under social motivation.

Form B of the Henmon-Nelson Test of Mental Ability for college students was administered to all patients during the same week that they were tested for reaction time. The patients were tested in small groups of less than 5 persons and aided in understanding the directions in order to maximize their performance on the test. The reaction-time and Henmon-Nelson tests were readministered two to four weeks from the original testing to all patients then available. Of the total sample, five of the privileged-young, two of the unprivileged-young, and one of the unprivileged-old patients were unavailable. Henmon-Nelson scores were available on the normal controls from a classroom administration; no retest was conducted.

### RESULTS

The effects of the three conditions of motivation upon reaction time in the schizophrenic and normal groups are presented in Table 1. Since a few extreme cases in the unprivileged-old schizophrenic group tended to influence the group means unduly, medians are also shown in the tables to permit an evaluation of the group trends relatively unaffected by extreme scores. Statistical analyses of the group differences have also taken the heterogeneity of this group into consideration, square root transformations of scores being employed where the group variances were not homogeneous. While the

TABLE 1  
REACTION TIMES IN .01 SEC. OF SCHIZOPHRENIC AND  
NORMAL GROUPS UNDER DIFFERENT CONDITIONS  
OF MOTIVATION

	Unprivileged Schizophrenics		Privileged Schizophrenics		Normals
	Old	Young	Old	Young	
Shock Motivation					
Mean	19.2	15.1	17.0	13.7	12.0
Mdn.	18.5	14.0	16.5	13.5	12.0
SD	6.2	4.2	4.4	2.3	1.6
Social Motivation					
Mean	146.8	38.3	35.6	23.1	18.8
Mdn.	84.5	34.5	33.0	22.5	18.0
SD	161.1	20.0	16.3	16.8	2.6
Anxiety Motivation					
Mean	124.7	51.5	73.5	32.7	24.8
Mdn.	96.0	45.0	60.0	32.5	24.5
SD	99.8	26.7	41.6	9.4	5.5

normal group was not comparable to the schizophrenic groups in age or sex composition, their average scores are shown in the tables, and comparisons have been made where it was intended to match schizophrenic Ss against an optimal level of normal functioning.

*Shock motivation.* The hypothesis that older schizophrenics show a reduced capacity to react rapidly to biological shock motivation appears to be clearly supported by the data shown in the first two rows of Table 1. Analysis of variance of these scores revealed that age was a significant source of variance at the .05 level, while privilege status did not significantly seem to affect reaction time. There was no significant interaction between age and privilege status. A separate *t* test showed that the normals were not significantly faster than the privileged-young patients.

*Social motivation.* The middle two rows of Table 1 show that the older patients and those without privileges reacted more slowly under social motivation. Analysis of variance indicated that both these sources of variance were significant at the .05 level; their interaction was not statistically significant.

Although these findings are not consistent with the hypothesis concerning the social motivation dimension, the data of both the present and the previous studies showed that the patients who could not react rapidly to shock were also unable to react rapidly to social instructions. Therefore, an index of social motivation was computed for every S which takes into account his biological capacity to react as indicated by reaction time under shock.

These social efficiency percentages, based on the ratio between speed under social motivation to speed under shock motivation, are presented in Table 2. The analysis of these data revealed that the privileged patients

TABLE 3  
HENMON-NELSON TEST SCORES OF SCHIZOPHRENIC AND NORMAL GROUPS

	Unprivileged Schizophrenics		Privileged Schizophrenics		Normals
	Old	Young	Old	Young	
Mean	7.2	16.6	14.0	27.5	47.1
Mdn.	6.0	18.0	12.5	22.5	46.5
SD	5.3	11.2	10.6	12.6	7.4

obtained significantly higher percentages than the unprivileged patients, but that age did not significantly affect these scores. All the normals and 8 of the 10 privileged-young patients showed greater than 50 per cent social efficiency: these two groups were not significantly different on this measure. These results are in line with the social motivation hypothesis.

*Anxiety motivation.* In keeping with the previous findings, it may be seen from the last two rows of Table 1 that the groups reacting slowly to shock and social motivation were also slowest under anxiety motivation. Age was a significant source of variance at the .01 level, while privilege status and the interaction were not significant. Furthermore, the privileged-young patients, who reacted at the normal level under shock and social motivation, were significantly slower under anxiety motivation than the normal group (.05 level).

*Intelligence test and correlation data.* The current mental functioning of the groups is indicated by the Henmon-Nelson scores, shown in Table 3. Since this form of the Henmon-Nelson test was standardized for college students and required concentration in following detailed clerical directions, the scores obtained by the patients might be expected to reflect both the social motivation to cooperate and the capacity to sustain attention. Analyses of these data revealed that youth, privilege status, and membership in the normal group were all significantly associated with higher scores.

The relationship of these mental efficiency scores to the reaction time indices of motivation is shown in the intercorrelation matrix presented as Table 4. The correlations indicate that Henmon-Nelson scores were significantly related to all three measures. Shock and anxiety reaction times were also significantly

TABLE 2  
SOCIAL EFFICIENCY PERCENTAGES OF SCHIZOPHRENIC AND NORMAL GROUPS

	Unprivileged Schizophrenics		Privileged Schizophrenics		Normals
	Old	Young	Old	Young	
Mean	31.2	45.6	56.3	60.0	63.0
Mdn.	26.0	45.0	49.0	59.5	63.5
SD	17.5	16.7	27.9	8.6	7.5



related; but social efficiency percentages were not significantly related to either of the other reaction time measures. Furthermore, it may be seen from this table that all these indices had useful retest reliabilities for measurement purposes; shock reaction time and Henmon-Nelson scores being highly stable.

In view of the fact that the Henmon-Nelson scores appeared to provide an index of general mental efficiency significantly related to the other measures, a multiple correlation coefficient was computed to determine the amount of variance in Henmon-Nelson scores that could be accounted for by all three reaction-time scores. This multiple  $R$  was equal to .57, accounting for 33 per cent of the variance in Henmon-Nelson scores. The multiple regression equation was:

$$Y' = 20.17 - .74X_1 + .24X_2 - .02X_3$$

where

$Y'$  = predicted Henmon-Nelson score

$X_1$  = shock reaction time

$X_2$  = social efficiency percentage

$X_3$  = anxiety reaction time

It is apparent from the regression weights and the beta coefficients obtained that most of the variance in mental efficiency in this sample of schizophrenics was attributable to the shock motivation score (12 per cent) and the social efficiency percentage (17 per cent). Anxiety reaction time, apart from that component related to shock reaction time, contributed less than four per cent to these schizophrenic mental efficiency scores, but

would probably contribute appreciably to the variance of a distribution composed only of early schizophrenic scores.

### DISCUSSION

The results of the present experiment suggest that three relatively separate dimensions of motivational deficit may be identified among the schizophrenics. There seems to be a class of schizophrenics who show a defective capacity to respond to biological stimulation in addition to the more typical reduction in social motivation. These patients tend to be older and to have poorer mental efficiency but, within the limits of their reduced biological reactivity, they are as socially efficient as other schizophrenics. Such patients may well show the quantitatively subnormal adrenal responses to stress observed by Hoagland (5) in chronic schizophrenics. It would also be profitable to investigate whether such psychotics display a significant number of organic signs on the Rorschach. An expectation of this sort is supported by the finding of Brackbill and Fine (2) that typical process schizophrenics are indistinguishable from organics in the incidence of Piotrowski's ten signs of organic involvement.

A second class of schizophrenics seems unimpaired in biological motivation, but the reactivity of these patients to social incentives is typically well under 50 per cent of their reactivity to shock. The extent of the deficit in social motivation appears unrelated to age but is reflected in a lack of ground privileges and poorer mental efficiency. This group illustrates the social disarticulation process described by Cameron (3) and others (12) in the typical disturbances in reaction time reported originally by the Worcester group (6, 8) and more recently by King (7). King, however, has rejected the motivational interpretation of slowed reaction times, suggesting rather that this symptom is "... a primary indication of disturbance in a basic adaptational process, that of motor adjustment to the external surround." The findings of both the present and previous papers that shock motivation eliminates this primary disturbance in motor adjustment, at least in some patients, makes his interpretation very questionable.

A third class of schizophrenics functions at

TABLE 4

RELIABILITIES AND INTERCORRELATIONS OF REACTION TIME AND HENMON-NELSON SCORES IN SCHIZOPHRENICS\*

	Shock Reaction Time	Social Efficiency Percentage	Anxiety Reaction Time	Henmon-Nelson Score
Shock reaction time	.94			
Social efficiency percentage	-.07	.79		
Anxiety reaction time	.64	-.26	.71	
Henmon-Nelson score	-.39	.44	-.39	.91

\* The reliabilities presented are Pearson  $r$ 's based on the scores of the 32 patients available for both the initial test and the retest. The remaining intercorrelations are Pearson  $r$ 's based on initial test scores of all 40 patients. For 38  $df$  a Pearson  $r$  of .31 is required for the .05 level of confidence.

a normal level of efficiency under both biological and social motivation. These patients are young, have ground privileges, and tend to display good mental efficiency. They seem to display abnormal reaction times only under the stress of anxiety motivation, suggesting a type of psychosis in which the disorder constitutes primarily a reaction to disruptive anxiety. While the intellectual regression of the first class of schizophrenics appears to resemble Beck's (1) reaction Form S-1, the intellectual control and lack of symptoms of this third class may well involve the defensive character hardening that he has described in his Form S-3.

The present findings do not indicate whether these three types may best be regarded as different forms of schizophrenia or as stages of the same schizophrenic process. However, the reaction-time technique employed appears to provide a brief, reliable instrument for assessing relatively fundamental pathology in motivation. It seems likely that prognosis is poorer in those patients showing a deficit in biological motivation, better in those with good biological motivation and high social efficiency percentages. Further exploration of the manner in which these dimensions of motivation are related to different prognostic and etiological factors should determine whether the types isolated require classification as separate forms of psychopathology or as various phases of a basic schizophrenic process.

#### SUMMARY

The present experiment was designed to investigate the relationship of age and ground privilege status to reaction time indices of biological, social, and anxiety motivation in schizophrenic patients.

Two conditions of both age and privilege status were arranged in a  $2 \times 2$  factorial design by subdividing 20 schizophrenics over 40 years and 20 under 40 years into two equal groups, each consisting of 10 privilege and 10 ward patients. Ten college students served as normal controls.

All Ss were tested on a reaction time task requiring them to lift a finger from a telegraph key in response to a buzzer during 80 successive trials, successively involving social instructions, unpredictable anxiety cues, and shock

stimulation. The Ss were also given the Henmon-Nelson test of Mental Ability and retested on both the reaction time and intelligence tests.

The results showed that age significantly slowed mean biological and anxiety reaction times but did not affect the social efficiency percentage based on the ratio of speed under social motivation to speed under shock motivation. Privilege status significantly improved mean social reaction time and social efficiency percentage but had no effect on the other two measures and did not interact with age. The normals were superior to the schizophrenics on all measures, but could be significantly differentiated from the younger, privileged group only on anxiety reaction time.

Correlational data were presented showing that shock reaction time was correlated with anxiety reaction time but independent of social efficiency percentage in the schizophrenic sample. Mental efficiency was significantly correlated with the three reaction time indices, and all four measures had stable retest reliabilities.

It was concluded that lessened reactivity to biological stress with increasing age, reduced responsiveness to social demands, and inability to control anxiety are three separate dimensions of motivational deficit which may be isolated in schizophrenic patients. It is suggested that these dimensions may be related to different etiological and prognostic factors, and may prove useful in the classification of schizophrenic pathology.

#### REFERENCES

1. BECK, S. J. *The six schizophrenias*. New York: American Orthopsychiatric Ass., 1954. Research Monograph No. 6.
2. BRACKBILL, G. A., & FINE, H. J. Schizophrenia and central nervous system pathology. *J. abnorm. soc. Psychol.*, 1956, **52**, 310-313.
3. CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 861-921.
4. HENMON, V. A. C., & NELSON, M. S. *The Henmon-Nelson tests of mental ability*. New York: Houghton-Mifflin, 1931.
5. HOAGLAND, H. Stress and the adrenal cortex with special reference to potassium metabolism. *Psychosom. Med.*, 1950, **12**, 142-148.
6. HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: Reaction time. *J. gen. Psychol.*, 1937, **16**, 39-82.



7. KING, H. E. *Psychomotor aspects of mental disease*. Cambridge: Harvard Univer. Press, 1954.
8. RODNICK, E. H., & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, **97**, 214-225.
9. ROSENBAUM, G., MACKAVEY, W. R., & GRISELL, J. L. Effects of biological and social motivation on schizophrenic reaction time. *J. abnorm. soc. Psychol.* 1957, **53**, 364-368.
10. SENF, RITA, HUSTON, P. E., & COHEN, B. D. Thinking deficit in schizophrenia and changes with amytal. *J. abnorm. soc. Psychol.*, 1955, **50**, 383-387.
11. SHAKOW, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. Dis. Monogr.*, 1946, No. 70.
12. SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington D. C.: W. A. White Foundation, 1947.

Received October 23, 1956.

# THE EFFECTS OF SEX-ROLE IDENTIFICATION UPON PROBLEM-SOLVING SKILL<sup>1</sup>

G. A. MILTON

*Stanford University*

**D**URING recent years, there has been an increased interest in the effects of "nonintellectual" processes upon problem solving and thinking. One area in which such effects can be and have been studied with readily discriminable criterion groups is the area of sex differences. Sweeney (5) has summarized several studies in which problem-solving differences between men and women were observed. He also reports experiments of his own which demonstrate that men solve certain classes of problems with greater facility than do women, even when differences in intellectual aptitude, academic training, and special abilities are controlled.

The present study explores the possibility that the differences in problem-solving skill between men and women may be due, at least in part, to a set of learned behaviors that constitute a culturally defined sex role. A simple extrapolation from the earlier studies of sex differences leads to the hypothesis that the more an individual identifies with the masculine sex role, the greater will be his problem-solving skill.

This hypothesis may be stated more explicitly: (a) There is a positive relationship between masculine identification and problem-solving achievement. (b) When an adjustment is made for the between-subjects variance contributed by sex-role identification, sex differences in problem-solving achievement will be reduced. The implication of the hypothesis is, of course, that not only will differences between men and women in problem-solving skill be reduced by an adjustment for masculine role identification, but also that the more "masculine" men will solve problems more readily than "feminine" men and, similarly, that the more masculine women

will be better problem solvers than women who have greater identification with the feminine role.

## METHOD

### *Indices of Sex-Role Identification*

The concept of sex role refers to a heterogeneous collection of actions that differentiate typically masculine behavior from typically feminine behavior in a particular culture. These characteristics of the sex role may be defined restrictively in terms of overt behavior only, or more broadly in terms of attitudes, interests, needs, and other inferred constructs in addition to overt behavior. The indices of sex role used in this investigation have been determined by primarily empirical means; overt and covert measures were used, the criterion for selection being that they involved nonintellectual factors that differentiate men from women for the general culture from which the experimental sample was drawn.

The *Terman-Miles M-F Test* (6) was chosen as the primary defining measure of sex-role identification because it measures a broader range of responses than any other sex-role measure cited in the literature. This test consists of a set of items that differentiate men from women in a culture similar to that of the current sample (cf. below). The test elicits responses that not only discriminate men from women, but also seem to assess a culturally defined masculinity-femininity variable within each sex.

Two secondary measures of sex role were also employed. These tests tap a narrower range of sex-role content. One, the *M-F scale of the Minnesota Multiphasic Personality Inventory* (2), consists largely of self-report items referring to feelings and emotions. The other, the *Behavioral Inventory M-F* (4) scale, is part of an unpublished experimental inventory that defines the role in terms of self-reported overt behavior that discriminates between men and women in a college population.

### *Problems*

Twenty problems that had shown significant sex differences in pretest trials were used. The problems were selected and refined by Nakamura (3), and his methods of administration were duplicated in the present study. Item analysis has indicated that the problems involve two types of problem-solving skill, restructuring and straightforward solution. The restructuring problems require the subject to alter his initial set in order to attain the solution; the straightforward problems are solvable by direct means. Ten straightforward and ten restructuring problems were arranged alternately by pairs, according to order of difficulty. One member of each pair involved a numerical solution and the other member was non-numerical.

<sup>1</sup> This experiment is one in a series of studies of problem-solving done under Project NR 150-149 and supported by Contract N6onr 25125 between Stanford University and the Office of Naval Research. Work on the contract was under the general direction of Dr. Donald W. Taylor. Permission is granted for reproduction, translation, publication, use, and disposal of this article in whole or in part by or for the United States Government.



The problems were bound in a booklet with one problem per page so that each problem was presented individually. A working time of four minutes per problem was allowed with a ten-minute intermission after the first ten problems. Problems were scored either correct or incorrect with no partial credits given.

### Subjects

One hundred and twenty-nine students from an introductory psychology class at Stanford participated as subjects. Of these, 63 were males and 66 were females. Scores on the mathematical and verbal sections of the College Entrance Examination Board Scholastic Aptitude Test were available for all subjects as a control upon intelligence.

### Procedure

Subjects were tested in six groups of 16 to 35 subjects. For all groups, the tests were administered in three sessions. The order of presentation of the tests was arbitrary. The MMPI M-F and the Behavioral scale M-F were presented during the first session; the problem-solving test during the second session, and the Terman-Miles M-F during the third session. Because of the nature of the items on the M-F scales, it was assumed that there would be no interaction with order of presentation; therefore, this arrangement was made purely on the basis of experimental expediency.

## RESULTS

The results of this investigation may be divided into two categories. First, sex differences on the problem-solving measure are examined in order to establish that those reported by other investigators (5) are found in the current investigation. The hypothesized relationship between sex role and problem solving is then tested. Both the relationships between sexes and within sexes are examined.

### Sex Differences in Problem Solving

The differences between the problem-solving achievement of men and women presented in Table 1 are the data which the hypothesis must partially explain. The differences between men and women are significant with respect to the total problem-solving score and all the sub-scores, but the difference between men and women on one subtest is not significantly greater than that on another subtest. The total score can therefore be used to test the effects of the sex-role measure. These results are similar to those obtained by Nakamura (3) for the same problem on a sample of University of California students.

TABLE 1  
SEX DIFFERENCES IN PROBLEM-SOLVING SKILL

Type of Problem	Men		Women		t
	Mean	SD	Mean	SD	
Problem-solving, total score	9.10	3.68	7.18	3.27	3.12**
Restructuring problems	4.33	2.22	3.38	1.88	2.63*
Straight-forward problems	4.76	2.17	3.80	1.79	2.74**
Numerical problems	5.06	2.27	4.04	1.87	2.78**
Non-numerical problems	4.05	2.15	3.14	1.79	2.59*

\* Significant beyond .02 level.

\*\* Significant beyond .01 level.

TABLE 2  
CORRELATION OF THE SEX-ROLE AND APTITUDE VARIABLES WITH PROBLEM-SOLVING SKILL

Variable	Men	Women	Combined
Terman-Miles M-F	.44**	.26*	.41**
MMPI M-F	.14	.13	.29*
Behavior M-F	.28*	.23	.36**
Verbal aptitude	.44**	.28*	
Math aptitude	.63**	.78**	

\* Significant beyond .05 level.

\*\* Significant beyond .01 level.

### Effects of Sex-Role upon Sex Differences in Problem Solving

In order to confirm the hypothesis of this investigation, two effects must be demonstrated: that there is a correlation between sex-role identification and problem solving, and that this correlation accounts for some of the differences between men and women in problem-solving skill. The correlations between sex-role measures and problem solving are presented in Table 2.

It is clear that problem-solving skill is significantly related to sex-role identification as measured by the Terman-Miles test. There is a significant relationship both within and across sexes. The evidence concerning the secondary measures of sex-role is less clear. The correlation of problem solving with the MMPI M-F scale approaches significance only for the combined sex group, and the correlation of problem solving with the Behavioral M-F scale does not appear to be significant for the women.

In order to determine if the differences between men and women in problem-solving score can be accounted for in terms of score on the measures of sex-role identification, a simple analysis of covariance was performed. The rationale of this operation is that if the

TABLE 3  
ANALYSIS OF COVARIANCE: SEX DIFFERENCES IN  
PROBLEM SOLVING WITH ADJUSTMENT FOR  
TERMAN-MILES M-F SCORE

Source of Variance	Before Adjustment		After Adjustment			
	df	Sum of Squares	df	Sum of Squares	Mean Squares	F
Total	128	1675	127	1375		
Within	127	1557	126	1366	10.84	
Between	1	118	1	9.32	9.32	0.86

TABLE 4  
ANALYSIS OF COVARIANCE: SEX DIFFERENCES IN  
PROBLEM SOLVING WITH ADJUSTMENT  
FOR MMPI M-F SCORE

Source of Variation	Before Adjustment		After adjustment			
	df	Sum of Squares	df	Sum of Squares	Mean Squares	F
Total	128	1675	127	1534		
Within	127	1557	126	1529	12.13	
Between	1	118	1	5.86	5.86	0.48

TABLE 5  
ANALYSIS OF COVARIANCE: SEX DIFFERENCES IN  
PROBLEM SOLVING WITH ADJUSTMENT FOR  
BEHAVIORAL M-F SCORE

	Before Adjustment		After Adjustment			
	df	Sum of Squares	df	Sum of Squares	Mean Squares	F
Total	128	1675	127	1462		
Within	127	1557	126	1454	11.54	
Between	1	118	1	7.76	7.76	0.67

correlation between the problem-solving score and the sex-role measure is sufficient to reduce significantly the difference in problem-solving score, then sex-role identification as measured by the particular test accounts for a part of this difference. Table 3 presents the results of this analysis for the Terman-Miles test, and Tables 4 and 5 show similar results for the MMPI and the Behavioral Inventory.

The results indicate that the Terman-Miles not only accounts for a significant part of the difference between men and women in problem-solving skill, but also that it diminishes this difference to the point where it is no longer significant. The results with the MMPI and Behavioral Inventory are comparable to those obtained with the Terman-Miles. The difference between men

TABLE 6  
BETA WEIGHTS: RELATIVE INDEPENDENT CONTRIBUTION OF EACH OF THE VARIABLES TO  
PROBLEM-SOLVING SCORE

Variable	Men (N = 63) Beta	Women (N = 66) Beta
Terman-Miles M-F	.297**	.183*
MMPI M-F	.074	.050
Behavior M-F	.066	.126
Verbal aptitude	.224**	.099
Math aptitude	.473**	.848**

\* Significant beyond .05 level.

\*\* Significant beyond .01 level.

and women on the problem-solving measure diminishes when the variance attributable to sex-role identification is controlled.

The hypothesis requires, by implication, that differences in problem-solving skill within a sex must also be partially accounted for by differences in sex-role identification. Table 2 indicates that the Terman-Miles measure was significantly correlated with problem solving within both sexes, but that the other sex-role measures were less clearly related to problem solving.

In order to determine what portion of the variance in the problem-solving task within a sex may be explained on the basis of sex-role identification and the other nonintellectual variables related to sex role, the variables were fitted to a multiple regression equation and the appropriate beta weights were determined for each variable by means of a Doolittle solution. Since the hypothesis requires that differences in problem-solving skill must be explained independently of differences in intellectual aptitude, the verbal and mathematical measures were included in the equation. The results of this analysis are presented in Table 6.

For both the men and the women, the Terman-Miles test measure of sex-role identification is a significant contributor to the variance in problem-solving scores, even when mathematical and verbal aptitude are held constant. The mathematical aptitude scale contributes a large portion of the variance for both sexes, and the verbal aptitude scale makes a significant contribution for the males. The other sex-role variables make no independent contribution to the variance in problem solving, when the more powerful Terman-Miles test is included in the battery.



## DISCUSSION

The results of this study seem to indicate that the differences between men and women in problem-solving skill as reported by a number of investigators are merely part of a more general difference between men and women in role identification. Men have characteristically different behavior, attitudes, emotions, and motivations than do women, and the differences in problem-solving skill reflect or are a part of these differences. To the extent that allowance can be made for these differences in general sex-role identification, the differences in problem-solving achievement seem to disappear.

Moreover, there are differences *within* a sex as to the extent of identification with the appropriate sex role, and these within-sex differences in role also seem to be related to problem-solving skill. Men who identify with the masculine sex role are better problem solvers than are those who identify with a more feminine role, and conversely, women who identify with the appropriate sex role have poorer problem-solving scores than do women who have a more masculine identification. All these comparisons, of course, assume a common level of intelligence.

This interpretation leads to several implications. For example, the acquisition of problem-solving skills may be affected by the same processes which govern the formation of sex-role identification. If the Freudian hypothesis concerning the identification process is correct, then important antecedents of adult problem-solving skill are to be found in childhood. The male child who forms an adequate identification with the male role, given sufficient intelligence and opportunity to learn, will grow up to be a good problem solver. On the other hand, the female child, even though possessing adequate intelligence and opportunity to learn, will probably not develop problem-solving skills (of the sort used in this study) if she forms an appropriate identification with the feminine role, because this type of problem solving is not appropriate to the female sex-role in her culture. This is, of course, a very broad extrapolation from the data of this experiment.

Another interesting implication concerns the possibility for improvement in problem-solving skill. If the sex-role interpretation of

the results of this experiment is correct, any attempt to change problem-solving skill will be more effective if corresponding changes are made in certain aspects of the conventional sex-role identification. While it may not be necessary to change the sex-role identification of a subject in order to change his (or her) problem-solving skill, it would seem necessary to neutralize the effects of sex-role identification that relate to the problem-solving process. Practice on problems alone probably would not be effective. Carey (1) has demonstrated an effect similar to that of partially changing the problem-solving attitudes associated with sex-role identification. Through group discussions of women's attitudes toward intellectual activity, she was able to improve their problem-solving scores. This change, though significant, was slight, and it would seem that more than a few discussions are required to change attitudes that appear to be basic to the culturally defined sex role. If one considers that sex-role identification has had an effect on the learning of a subject for the major part of a lifetime, it seems that a change in sex role would require a long period of relearning. Also, if major changes in problem-solving performance are to be achieved, the relearning of sex-role would necessarily be accompanied by a period of relearning of problem-solving skills.

Although this interpretation of the results seems appropriate for a college population, there are logical difficulties in any extrapolation of these results to the more general population. Wherever there are subsections of the general population in which problem solving, as defined in this investigation, is not appropriate to the masculine role, this relationship would not be expected. Since the studies of sex differences in problem solving which were mentioned earlier have been conducted upon college populations, there is little evidence as to how far these conclusions can be generalized.

## SUMMARY

The hypothesis of the study was that sex differences in problem-solving skill may be partially accounted for by differences in sex-role identification. The Terman-Miles M-F Test was used as the primary index of sex-role identification. Two other M-F ques-

tionnaires were also employed. These scales and a measure of problem-solving skill were administered to 63 male and 66 female university undergraduates. In general, the results indicate that there is a positive relationship between masculine sex-role identification and problem-solving skill both across sexes and within a sex. When allowance is made for this relationship, the difference between men and women in problem-solving performance is diminished.

#### REFERENCES

1. CAREY, G. L. Reduction of sex differences in problem solving by improvement of attitude through group discussion. Dept. Psychol., Stanford Univer., 1955, (ONR Tech. Rep. No. 9, Contract N6onr 25125).
2. HATHAWAY, S. R., & MCKINLEY, J. C. *Minnesota Multiphasic Personality Inventory*. New York: Psychol. Corp., 1951.
3. NAKAMURA, C. Y. The relation between conformity and problem solving. Dept. Psychol., Stanford Univer., 1955. (ONR Tech. Rep. No. 11, Contract N6onr 25125).
4. PINNEAU, S. R., & MILTON, G. A. The objective measurement of overt behavior. Paper read at Western Psychol. Ass., June, 1954.
5. SWEENEY, E. J. Sex differences in problem solving. Dept. Psychol., Stanford Univer., 1953, (ONR Tech. Rep. No. 1 Contract N6onr 25125).
6. TERMAN, L. M., & MILES, C. C. *Sex and personality*. New York: McGraw-Hill, 1936.

Received October 24, 1956.



# THE EFFECTS OF CLEAR AND UNCLEAR ROLE EXPECTATIONS ON GROUP PRODUCTIVITY AND DEFENSIVENESS<sup>1</sup>

EWART E. SMITH<sup>2</sup>

*University of Colorado*

**I**N ALL groups, people invest considerable time and energy trying to develop sufficient understanding of other members in order to feel they can predict how these other members will behave. In role theory concepts (5), this is attempting to locate the position of the other so that one may have valid role expectations of the other's role enactments. The same phenomena have been discussed in other theoretical terms. Bales (1) states that the social structure of groups develops in order to "... reduce the tensions growing out of uncertainty and unpredictability in the actions of others."

The general hypothesis of the present study is that the time and energy used by group members in attempting to predict the behavior of others reduces the amount of group energy available for any given group goal or task. This hypothesis is consistent with the conception of group energy formulated by Cattell (2), who states that group members have a finite amount of energy available in a group situation, and that the amount of energy absorbed by the internal group structure is subtracted from the amount available for group productivity.

The present investigation is primarily concerned with the relationship between the degree to which members of a small group feel they can predict each other's behavior, and the group's task efficiency. In addition to the positive relationship anticipated between these variables, it was expected that uncertainty regarding the roles of the other group members would be reflected in increased defensiveness and dissatisfaction with the group experience.

<sup>1</sup> Condensed from a dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Department of Psychology, University of Colorado. The writer wishes to express his appreciation to Professor Jack R. Gibb, under whom the work was done, and to Professors Maurice P. Smith and Michael Wertheimer for their helpful criticisms and advice. The work was supported by a research grant (Contract Nonr-1147(03), NR 170-226) from the Group Psychology Branch of the Office of Naval Research.

<sup>2</sup> Now with the RAND Corporation.

While these problems have not previously been studied experimentally, as far as the author can ascertain, Torrance (8) developed similar views from his study of 1,000 critical incidents of Air Force crews downed in enemy territory. He concluded that the two major detrimental factors in group survival were an unclear situation and an unclear group structure. Regarding the latter factor, he states, in part:

Under stress, these linkages between members may become confused and thus people do not have a clear perception of what they can expect from one another, with whom they can relate, how they can relate to one another, and so on. The fact that this happens apparently influences the ability of the group to survive.

In order to induce ambiguous role expectations, paid accomplices were assigned roles as silent members in small groups of persons who were, initially, strangers to each other. In the initial meeting of such a group, each member should find the behavior of all of the others equally unpredictable. But in the course of group interaction the behavior of those who spoke should become more predictable, while the behavior of those who remained silent should remain unpredictable. Eventually the group members would undoubtedly come to predict with confidence the simple fact that the silent members would remain silent. Since only short-term groups were studied, however, the latter effect is not relevant to the present experiment.

This reasoning was substantiated in several preliminary investigations. The presence of silent group members in short-term groups had little, if any, effect on productivity during the first few minutes, but productivity decreased to a low point at about thirty minutes, followed by a gain until the presence of silent members resulted in no decrement in productivity.

The hypotheses tested in the present study are:

1. The presence of silent members in a group produces (a) decreased productivity, (b) increased defensiveness, and (c) increased dissatisfaction with the group experience.

2. Decreasing the ambiguity of role expectations in a group that contains silent members counteracts the effects predicted in Hypothesis 1.

### METHOD

#### Subjects and Task

The subjects (Ss) were 50 males and 90 females from the introductory psychology classes at the University of Colorado. The task, taken from an experiment by Taylor and Faust (7), is an adaptation of the parlor game, "Twenty Questions," in which the Ss had to identify items designated by the experimenter (E) as either animal, vegetable, or mineral. Each problem had to be solved with less than forty questions for the group to receive credit. There was no time limit on individual problems. To each question asked by a group member, E replied in one of the following ways:

(a) Yes; (b) No; (c) Partly; (d) Sometimes; (e) Not in the usual sense of the word; (f) I don't know (no charge for the question); (g) Please restate the question (if the question was unclear or could not be answered in one of the above ways). Difficult items like wrench, ruby, and bread were used.

#### Experimental Design

In the experimental Conditions A and B, each group was composed of three participating Ss, two females and one male, and two silent members, one female and one male. Condition A was designed to induce ambiguous role expectations. Condition B differed from Condition A in that the roles of the silent members were clarified prior to experimentation; all group members checked their "usual role" in a group on a check list. Everyone then announced the role he had checked, which was then checked by E on a large replica of the "Usual Roles" check list. The silent members always checked the "Listener" role.

In the control Conditions C and D, there were no paid silent members. In Condition C each group was composed of three participating Ss, two females and one male. In Condition D, each group was composed of five participating members, three females and two males.

In all conditions, Ss worked on the task for two periods. The first work period, which lasted fifteen minutes, was intended to allow ambiguous role expectations of the silent members to develop. In the second work period, which was ten minutes long, group productivity was measured. The problems, and their order, were the same for all groups. Regardless of how many problems the group had solved in the first work period, each group started with a new, and identical, set of problems in the second work period.

The thirteen silent paid participants were not known to the Ss in their groups. Each silent member was assigned to Condition A and B groups an equal number of times.

#### Dependent Variable Measurements

Group productivity was measured by the number of problems solved. *Defensiveness* was measured by a 25-item Likert scale, with a corrected, split-half reli-

TABLE 1  
MEDIAN RATING OF Ss AND SILENT PAID PARTICIPANTS ON ROLE AMBIGUITY

Statistics on Role Ambiguity Rating	Data under each condition					
	A Ambiguity (Silent Members)		B Clarity (Silent Members)		C Control (N = 3)	D Control (N = 5)
	Rating of Ss	Rating of Silent Members	Rating of Ss	Rating of Silent Members	Rating of Ss	Rating of Ss
Median	11.00	6.00	11.00	8.00	8.50	8.00
N	29	29	30	30	30	50

Note.—Low scores indicate role ambiguity.

H for these data is 34.89,  $p < .01$ .

N is 29 in Condition A as one S failed to fill out the rating form.

ability of .79, which had been validated by the known group method (4). A typical item is "I felt free to express all my ideas in this group." *Group satisfaction* was measured by a 15-item Likert scale, with a corrected split-half reliability of .93, which had been validated by the known group method. A typical item is "This is the most stimulating experience I've had in a long time." A single item was used as an index of *role ambiguity*: "I can predict how this person would act in other situations." The Ss rated each group member on a scale from -3 to +3. A numbering system was used so that Ss did not give their names. This measure was used as a check on the effectiveness of the experimental manipulation.

### RESULTS AND DISCUSSION<sup>1</sup>

Table 1 contains the findings on the ambiguity item, "I can predict how this person would act in other situations." Plotting the scores on this item revealed that normal distributions could not be assumed. Consequently, nonparametric rank tests have been used (3, 6). An *H* test, which is functionally equivalent to a one-way analysis of variance, has been computed on the data in Table 1. As this *H* is statistically significant, additional rank tests have been computed. Using this method, *z*'s are obtained which are normal deviates evaluated by the use of the normal probability table. The *z* (for related measures) on the ratings of silent members compared to the ratings of Ss in the Condition A groups is 4.46\*\*, indicating that the independent variable manipulation in Condition A was successful; i.e., that the silent paid participants were seen as less predictable than the Ss.

<sup>1</sup> Statistical significance at the .05 level has been indicated by a single asterisk; at the .01 level or better by a double asterisk.



Additional support for the hypothesis that silent members were seen as less predictable is given by the significant  $z$ , 4.37\*\*, between the ratings of silent members and  $S_s$  in the Condition B groups. However, comparison of the ratings of the silent members in Condition A with the ratings of the silent members in Condition B indicates that the difference, although in the predicted direction, is not significant. Therefore, the hypothesis that role clarification reduces the ambiguity of the roles of silent members is not supported.

This failure may be due to a temporal factor. At the time that the  $S_s$  rated the silent members on predictability, they had a period of 30 minutes of observations upon which to base their ratings. This 30-min. time sample was the same in Conditions A and B, and was apparently more powerful than the earlier one-minute role clarification in Condition B, perhaps resulting in a leveling effect on the comparison of Conditions A and B with respect to the ratings of the silent members. Results more in line with the hypothesis might have been obtained if the rating item had been given earlier.

An interesting, and unpredicted, finding is that the mean rating of the  $S_s$  in both Conditions A and B is significantly higher than the ratings of  $S_s$  in Condition C, with  $z$ 's of 3.24\*\* and 3.64\*\*, using a rank test for independent samples (3). This finding cannot be explained from the data, but, as a possible explanation, it may be suggested that in the groups containing silent members, the other  $S_s$  were seen by contrast as much more predictable than they otherwise would have been.

Table 2 contains the findings on group productivity. Plotting the group task scores for the second work period (as anticipated, there were no significant differences in the first work period) revealed that normal distributions could not be assumed. Consequently, nonparametric rank tests (for independent measures) were used. The obtained  $z$  between Conditions A and B is 2.18\*; between A and C, 2.35\*; and between A and D, 2.63\*\*.

Table 2 shows that the silent members reduced group productivity significantly in Condition A. These results support the hypothesis that ambiguous role expectations reduce group productivity. At first glance, some more parsimonious explanations might

TABLE 2  
NUMBER OF PROBLEMS SOLVED, UNDER EACH OF  
FOUR CONDITIONS, IN FIRST AND SECOND  
WORK PERIODS

Statistic on Number of Problems Solved	Data Under each Condition			
	A Ambiguity (Silent Mem- bers)	B Clarity (Silent Mem- bers)	C Control ( $N = 3$ )	D Control ( $N = 5$ )
Median: First work period (15 minutes)	5.50	6.50	6.00	6.00
Median: Second work period (10 minutes)	0.	3.00	2.00	2.50
Number of groups	10	10	10	10

Note.— $T$  for these data (second work period) is 9.29,  $p < .05$ .

TABLE 3  
AVERAGE DEFENSIVENESS SCALE SCORES UNDER EACH  
OF FOUR CONDITIONS

Statistic on D-Scale Scores	Data Under each Condition			
	A Ambiguity (Silent Members)	B Clarity (Silent Members)	C Control ( $N = 3$ )	D Control ( $N = 5$ )
Mean	66.00	61.53	49.17	52.74
SD	13.92	12.41	14.14	14.10
$N$	30	30	30	50

Note.—High scores indicate defensiveness.

The analysis of variance is not summarized, to conserve space, as only predicted differences have been tested. However, the overall  $F$  on the means is 9.72,  $p < .01$ .

be suggested. Thus, one might suppose that the decrement in productivity in Condition A was produced by silent members per se, rather than by ambiguous role expectations. However, silent members were also present in the Condition B groups, where they were as silent, "uncooperative," "unproductive," etc., as they were in the Condition A groups. The difference was that in the Condition B groups, the  $S_s$  expected the silent members to be silent. Therefore, the variable to which the decrement in productivity must be attributed was the unpredictability of the silent members.

The data in Table 3 support the first hypothesis, that the presence of silent members produces increased defensiveness. The  $t$  ratios between Conditions A and C, and A and D, are 4.48\*\* and 4.04\*\*, respectively. The second hypothesis, that the presence of silent members does not produce increased defensiveness if the roles of the group members are explicit, is not supported. The differences

TABLE 4  
AVERAGE GROUP SATISFACTION SCALE SCORES UNDER  
EACH OF FOUR CONDITIONS

Statistic on GS-Scale Scores	Data Under each Condition			
	A Ambiguity (Silent Members)	B Clarity (Silent Members)	C Control (N = 3)	D Control (N = 5)
Mean	54.80	62.60	74.77	74.50
SD	9.74	13.82	14.81	13.06
N	30	30	30	50

Note.—High scores indicate satisfaction.

The  $F$  ratio between the variances of Conditions A and C is 2.31; the over-all  $F$  on the means is 18.34,  $p < .01$  in both instances.

between Condition B and Conditions C and D on defensiveness are significant, with  $t$ 's of 3.45\*\* and 2.79\*\*, respectively. The difference between Conditions A and B is not significant.

The data in Table 4 clearly support the first hypothesis, that the presence of silent members reduces group satisfaction. The  $t$  between Conditions A and D is 9.26\*\*, and that between Conditions A and C is 6.05\*\*. As the  $F$  ratio on the variances of Conditions A and C is 2.31\*, a correction for  $t$ 's computed when variances are heterogeneous has been used (3). The second hypothesis, that the presence of silent members does not reduce group satisfaction if the roles of the group members are explicit, is partially supported by the data. Group satisfaction is higher in Condition B than in Condition A, with a  $t$  of 2.48\*. However, the decrement in satisfaction produced by the silent members was not completely offset by the role clarification in Condition B groups, as the  $t$  ratios between Condition B, and Conditions C and D, on group satisfaction, are 3.23\*\* and 3.81\*\*, respectively.

At this point, it should be noted that although the primary independent variable, the presence of silent members, produced the expected results in all cases, the lack of significant differences between Conditions A and B on the role ambiguity item and on the Defensiveness scale throws some doubt on the interpretation of the results given here. One must decide whether Conditions A and B were really different, as indicated by the significant differences between them on productivity and the Group Satisfaction scale; or whether they are really similar, as indicated

by the lack of significant differences on the ambiguity item and Defensiveness scale.

The argument here is that Conditions A and B were different, and that role ambiguity and role clarification were the variables which lowered and raised group productivity. The lack of significant differences between Conditions A and B on the role ambiguity item and the Defensiveness scale are interpreted as a product of their relatively low reliability and of the decay of the role clarification set-variable by the time the questionnaires were given. However, the data are open to alternate interpretation.

### SUMMARY

In a laboratory investigation, paid participants, who were instructed to remain silent, were included in small groups to induce ambiguous role expectations. Five-member problem-solving groups, each including three naive Ss and two silent members, were used in each experimental condition. Three- and five-member groups were used in control conditions. It was hypothesized that ambiguous role expectations would reduce group productivity and satisfaction, and increase defensiveness.

All these predictions were well supported by the data.

Silent paid participants were included in another condition, where group members revealed their intended role behaviors. Group members thus had accurate expectations that the paid participants would remain silent. It was hypothesized that, as a result of this role clarification, the effects predicted above would be reduced or negated. Thus, the presence of silent participants would not reduce group productivity and satisfaction, or increase defensiveness.

Role clarification, with silent paid participants present in the groups, did, as predicted, result in normal (i.e., control group level) group productivity. Group satisfaction was significantly increased, but not to the normal level. The prediction that role clarification would restore the decrement in defensiveness was not supported.

### REFERENCES

1. BALES, R. F. A theoretical framework for interaction process analysis. In D. Cartwright and



- A. F. Zander (Eds.), *Group dynamics: Research and theory*. Evanston, Ill.: Row, Peterson, 1953.
2. CATTELL, R. B. New concepts for measuring leadership, in terms of group syntality. *Hum. Relat.*, 1951, 4, 161-184.
  3. EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rhinehart & Co., 1954.
  4. GIBB, J. R. *Factors producing defensive behavior within groups*. Quarterly Status Report, Office of Naval Res., Contract Nonr-1147(03), NR 170-226, Feb. 15, 1956, 5 pp.
  5. SARBIN, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
  6. SIEGEL, S. *Non-parametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
  7. TAYLOR, D. W., & FAUST, W. L. Twenty questions: Efficiency in problem solving as a function of size of group. *J. exp. Psychol.*, 1952, 44, 360-368.
  8. TORRANCE, E. P. The behavior of small groups under the stress of conditions of "survival." *Amer. sociol. Rev.*, 1954, 19, 751-755.

Received October 24, 1956.

# THE RELATIONSHIP BETWEEN OVERT AND FANTASY AGGRESSION AS A FUNCTION OF MATERNAL RESPONSE TO AGGRESSION

GERALD S. LESSER

*Hunter College*

**I**N RECENT years, a voluminous literature has developed around the problem of establishing relationships between fantasy behavior and overt behavior. Different researchers have used different drive areas, different populations, different theoretical bases, and different methods of measurement. The most conspicuous conclusion is that the empirical findings are not in agreement.

The importance of this area of investigation for both clinical practice and personality theory has been elaborated by Lindzey (12). He concludes that one of the most important and difficult problems is the "determination of the conditions under which inferences based upon projective material directly relate to overt behavior and the conditions for the reverse" (12, p. 18). The present study concerns the differential conditions under which aggressive behavior is learned that may allow prediction of how aggressive expressions in fantasy are related to those in overt behavior.

Various studies (9, 13, 16, 18, 19) have demonstrated that the degree of correspondence between fantasy behavior and the associated overt behavior is greater for certain drives than for others. Significant positive correlations have been reported between TAT fantasy and overt behavior for variables such as abasement, achievement, creation, dependence, exposition, nurturance, etc. Significant negative correlations have been reported for sex, and inconclusive results have been obtained for a wide variety of other variables. For the variable of aggression, results include significant positive correlations between fantasy and overt expressions (8, 14), significant negative correlations (5, 16), and inconclusive findings (1, 2, 3, 4, 6, 10, 13, 15, 16, 18).

To resolve these inconsistent results, it has been suggested (13, 14, 16, 19) that motives that are culturally encouraged are "... likely to be as strong in their overt as in their covert manifestations" (13, p. 16), while motives that are culturally discouraged

are apt to show little or no relationship between the strength of fantasy and overt expressions.

Mussen and Naylor (14) have attempted to test the first segment of this formulation. They contended that lower-class culture encourages aggression, and predicted that "... in a lower-class group, individuals who give evidence of a great deal of fantasy aggression will also manifest more overt aggression than those who show little aggression in their fantasies" (p. 235). A mixed group of white and Negro boys, "... almost all of whom had been referred to the Bureau of Juvenile Research for behaviors which brought them into conflict with school and court authorities..." (p. 236), were used as subjects. The authors report a statistically significant but not especially strong positive relationship between ratings of overt aggression and number of aggressive TAT themes. Further investigation of Mussen and Naylor's hypothesis would profit from more precise measurement of parental response to aggression, control comparisons, and a more representative sample.

The present study seeks to examine the comparative consequences of both encouragement and discouragement of aggression through the hypothesis that under conditions of maternal encouragement of aggression a greater degree of correspondence exists between fantasy and overt aggression of children than under conditions of maternal discouragement of aggression.

## METHOD

### *Subjects*

The subjects (Ss) were 44 white boys (ages 10-0 to 13-2) and their mothers. The boys were drawn from one fifth grade and two sixth grades in two public schools. All of the boys and their mothers in these three classes participated except one mother who refused to be interviewed. The Kuhlmann-Anderson intelligence quotients of the boys ranged from 82 to 119, with a mean of 102. The two schools are in adjacent districts and the families constitute a relatively homogeneous upper lower-class group.



### *Maternal Attitudes and Practices*

Only one aspect of the environmental conditions of learning of aggressive behavior was measured, i.e., the maternal attitudes and practices supporting or prohibiting aggression. A structured questionnaire-interview schedule was orally administered to the mothers in their homes by a male interviewer. Questions regarding the support or prohibition of aggression constituted only one segment of the total interview; the entire interview schedule is described in detail elsewhere (11). Pertinent to the present study were eight items concerning the mother's attitudes toward aggression in children, and thirteen items about the mother's practices in dealing with the aggressive behavior of her child. An illustrative item measuring maternal attitudes toward aggression is: "A child should be taught to stand up and fight for his rights in his contacts with other children." The four response alternatives of agree, mildly agree, mildly disagree, and disagree were allowed for this item. An example of an item measuring maternal practices concerning aggression is: "If your son comes to tell you that he is being picked on by a bully at the playground who is his own age and size, there would be a number of different things you might tell him. Would you tell him to ignore him and turn the other cheek?" Response alternatives for this item were yes and no. Items that did not involve judgments on a four-point scale were transformed to have approximately the same range of scores as the items that involved four alternatives.

A single score was obtained for each mother by combining all items, assigning plus scores to the responses indicating support of aggression and minus scores to responses indicating discouragement of aggression. The range of scores was from +9 to -7, with a median score of +2. The corrected odd-even reliability coefficient was .80.

The distribution of scores for maternal response to aggression was dichotomized to form one group of mothers (with scores above or at the median) whose attitudes and practices were more supportive of aggressive behavior than those of the other group (with scores below the median). The hypothesis demands that the correlation between fantasy and overt aggression for the children of the mothers in the former group be significantly more positive than the corresponding correlation for the children of the mothers in the latter group.

### *Fantasy Aggression*

Fantasy aggression in the children was measured through an adaptation of the TAT procedure (13, p. 3-5). A set of ten pictures was designed. In each picture two boys are interacting. The pictures differed from one another in the degree to which the instigation to aggression was apparent.

To insure complete and accurate transcription of the stories, tape recordings were taken. An introductory period preceding the fantasy task served both to establish rapport between the child and the male examiner, and to familiarize the child with the recording device. Instructions were:

I'm going to show you some pictures. These are pictures of two boys doing different things. What

I'd like you to do is make up a story to each of these pictures. You can make up any story you wish; there are no right or wrong stories. Say what the boys are thinking and feeling and how the story will turn out.

The ten pictures, in the order of presentation, were:

1. One boy is holding a basketball and the other boy is approaching him with arms outstretched.
2. One boy is stamping upon an ambiguous object and the other boy is reaching for the object.
3. One boy is sitting behind the other boy in a classroom and is leaning toward him.
4. One boy is walking down the street and the other boy, with fists clenched, is glaring at him.
5. One boy, with fists clenched, is staring at the other boy who is sitting, head bowed, on a box.
6. One boy is sawing a piece of wood and the other boy is leaning on a fence between them, talking to him.
7. The two boys, surrounded by a group of other boys, are approaching each other with arms upraised and fists clenched.
8. The two boys are making a fire. One boy is kneeling to arrange the wood and the other boy is approaching, laden with wood for the fire.
9. One boy, who is looking back, is running down a street and the other boy is running behind him.
10. Two boys are standing in a field. One boy, with his hand on the other boy's shoulder, is pointing off in the distance.

A fantasy aggression score was obtained for each S by counting the number of times the following acts appeared in his stories: fighting, injuring, killing, attacking, assaulting, torturing, bullying, getting angry, hating, breaking, smashing, burning, destroying, scolding, expressing contempt, expressing disdain, cursing, swearing, threatening, insulting, belittling, repudiating, ridiculing.

Fantasy aggression scores ranged from 1 to 15, with a mean of 5.3. The corrected matched-half reliability coefficient was .86; the inter-judge scoring reliability coefficient was .92.

### *Overt Aggression*

To measure overt aggression in the child, a modified sociometric device, the "Guess who" technique (7), was adopted. The Ss were presented with a booklet containing a series of written descriptions of children, and asked to identify each of these descriptive characterizations by naming one or more classmates. Fifteen overt aggression items were used, such as "Here is someone who is always looking for a fight." A diversity of aggressive behaviors were included; items depicted verbal, unprovoked physical, provoked physical, outburst, and indirect forms of aggressive behavior.

An overt aggression score was obtained for each subject by counting the number of times he was named by his classmates. There were substantial differences among the three classes in the distributions of the overt aggression scores; in order to combine into one distribution the scores of children in different classes, overt aggression raw scores were transformed into standard scores.

The biserial correlation coefficient between the overt aggression measure derived from the children

and teacher entries for the same "Guess who" aggression items was .76 ( $p < .01$ ).

### RESULTS

Two Pearson product-moment correlation coefficients were obtained. For boys ( $N = 23$ ) whose mothers are relatively encouraging or supportive of aggression, the correlation between fantasy aggression and overt aggression is  $+.43$  ( $p < .05$ , two-tailed test). For boys ( $N = 21$ ) whose mothers are relatively discouraging of aggression, the corresponding correlation is  $-.41$  ( $p < .10$ , two-tailed test). These coefficients are statistically different ( $p = .006$ , two-tailed test).

When the total sample is not separated into two groups on the basis of scores for maternal response to aggression, the overall Pearson product-moment correlation coefficient is  $+.07$ . This coefficient is not significantly different from zero.

### DISCUSSION

Confirmation is found for the hypothesis that under conditions of relative maternal encouragement of aggression, a greater degree of correspondence exists between the fantasy and overt aggression of children than under conditions of relative maternal discouragement of aggression. Thus, the direction and extent of the relationship between fantasy and overt aggression in the child is apparently influenced by the maternal attitudes and practices surrounding the learning of aggressive behavior.

It has been predicted (13, 16) that those tendencies which are negatively sanctioned or prohibited will be high in fantasy expression and low in overt expression. This association is premised upon a compensatory or substitutive role of fantasy where overt expression is not allowed. A scatter plot of the fantasy and overt aggression scores for the children whose mothers discourage aggression (from which the  $-.41$  coefficient is derived) reveals a considerable number of such high fantasy aggression, low overt aggression scores. However, children with low fantasy aggression and high overt aggression scores are as well represented in this scatter plot as those with high fantasy aggression, low overt aggression scores. Although mothers of children in this group were classified (relative to the others) as discouraging aggression, perhaps certain

of them do so ineffectively, and thus allow the child sufficient release of aggressive feelings in overt behavior so that he may not need to express aggression in fantasy. An alternative speculation regarding the concurrence of low fantasy aggression and high overt aggression in the group exposed to maternal discouragement of aggression suggests that a child with strong aggressive needs whose mother prohibits aggression may assign this prohibitory attitude to the adult experimenter and suppress fantasy aggression expressions in the testing situation; yet this child may find avenues for overt expression of aggression among his peers.

In the present study, only one condition related to the learning of aggressive responses and controls was assessed, maternal attitudes and practices. Other possibly critical determinants that remain to be explored include fathers' behavior and teachers' attitudes and practices. This study has sampled a limited range of maternal attitudes and practices concerning aggression. Although there is no direct manner of determining the absolute degree of punitiveness of the most prohibitive mother in this sample, it appears unlikely that extremely severe and continuous maternal punitiveness is represented. Such severe condemnation of aggression might so limit or restrict both the fantasy aggression and overt aggression expressions of the child that no correlational analysis within such a group would be possible. Both the extremes of unimpeded permissiveness and severe condemnation warrant further investigation.

### SUMMARY

The relationship between fantasy and overt expressions of aggression was studied as a function of the maternal attitudes and practices toward aggression. Subjects were 44 boys and their mothers. The boys' fantasy aggression was assessed through a modified TAT approach, their overt aggression was measured through a modified sociometric technique, and maternal attitudes and practices toward aggression were measured by use of a questionnaire-interview device.

Support was found for the hypothesis that under conditions of maternal encouragement of aggression, a greater degree of correspondence exists between fantasy and overt aggression



sion of children than under conditions of maternal discouragement of aggression.

## REFERENCES

1. BACH, G. R. Young children's play fantasies. *Psychol. Monogr.*, 1945, 59, No. 2 (Whole No. 272).
2. BIALICK, I. The relationship between reactions to authority figures on the TAT and overt behavior in an authority situation by hospital patients. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1951.
3. CHILD, I. L., FRANK, KITTY F., & STORM, T. Self-ratings and TAT: Their relations to each other and to childhood background. *J. Pers.*, 1956, 25, 98-114.
4. DAVIDS, A., HENRY, A. F., MCARTHUR, C. C., & McNAMARA, L. F. Projection, self evaluation, and clinical evaluation of aggression. *J. consult. Psychol.*, 1955, 19, 437-440.
5. FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, 50, 3-11.
6. GLUCK, M. R. The relationship between hostility in the TAT and behavioral hostility. *J. proj. Tech.*, 1955, 19, 21-26.
7. HARTSHORNE, H., & MAY, M. A. *Studies in the nature of character. II. Studies in service and self-control*. New York: MacMillan, 1929.
8. KAGAN, J. The measurement of overt aggression from fantasy. *J. abnorm. soc. Psychol.*, 1956, 52, 390-393.
9. KAGAN, J., & MUSSEN, P. H. Dependency themes on the TAT and group conformity. *J. consult. Psychol.*, 1956, 20, 29-33.
10. KORNER, ANNELIESE F. *Some aspects of hostility in young children*. New York: Grune & Stratton, 1949.
11. LESSER, G. S. Maternal attitudes and practices and the aggressive behavior of children. Unpublished doctor's dissertation, Yale Univer., 1952.
12. LINDZEY, G. Thematic Apperception Test: Interpretive assumptions and related empirical evidence. *Psychol. Bull.*, 1952, 49, 1-25.
13. MURRAY, H. A. *Thematic Apperception Test Manual*. Cambridge: Harvard Univer. Press, 1943.
14. MUSSEN, P. H., & NAYLOR, H. K. The relationships between overt and fantasy aggression. *J. abnorm. soc. Psychol.*, 1954, 49, 235-240.
15. PITTLUCK, PATRICIA. The relation between aggressive fantasy and overt behavior. Unpublished doctor's dissertation, Yale Univer., 1950.
16. SANFORD, R. N., ADKINS, MARGARET M., MILLER, R. B., COBB, E. A., *et al.* Physique, personality, and scholarship: A cooperative study of school children. *Monogr. Soc. Res. Child Developm.*, 1943, 8, No. 1.
17. SEARS, R. R. Relation of fantasy aggression to interpersonal aggression. *Child Developm.*, 1950, 21, 5-6.
18. SYMONDS, P. M. *Adolescent fantasy: An investigation of the picture story method of personality study*. New York: Columbia Univer. Press, 1949.
19. TOMKINS, S. S. *The Thematic Apperception Test*. New York: Grune & Stratton, 1947.

Received November 2, 1956.

# AFFILIATION MOTIVATION AND PRODUCTIVITY IN SMALL GROUPS

RICHARD DECHARMS<sup>1</sup>

*University of North Carolina*

THE general hypothesis underlying the research reported here is that there is a systematic relationship between the strength of affiliation motivation and cooperative group behavior. At first glance one might expect that affiliation motivation would facilitate cooperative group behavior and inhibit competitive behavior. It seems plausible that a person high in the affiliation motive should perceive working for a group goal as an activity instrumental to acceptance by the group, and therefore be predisposed to cooperate rather than compete.

These simple predictions are complicated by at least two factors. First, it has been shown (1, 5) that the experimental arousal of motives depends on the conditions under which *S* works. It follows that the predictions should hold only under certain arousal conditions. A second factor has been pointed out by French and Chadwick (2) who found that affiliation motivation can be predominantly goal-oriented or predominantly threat-oriented, and that these different orientations may have different behavioral concomitants. For instance, unpopular *Ss* tend to show more threat-oriented responses on a projective measure of motivation.

Threat-oriented and goal-oriented motivation probably result from differential experiences of acceptance in the past. The goal-oriented person reintegrates positive past experiences of acceptance in an affiliation situation, while the threat-oriented person reintegrates negative experiences of rejection. The person who has been accepted in the past may have learned that cooperation

leads to further acceptance. The previously rejected person, on the other hand, may desire acceptance so strongly that anticipation of rejection is particularly frustrating and leads to an aggressive competitive response. Affiliation motivation of the threat-oriented type may therefore facilitate competition rather than cooperation. Competition may in turn have been rewarded by recognition, which may even have led to greater acceptance of the originally rejected person.

These considerations lead to the following hypotheses:

*Hypothesis I.* Under conditions that lead them to anticipate rejection, *Ss* high in threat-oriented affiliation motivation are more productive on a competitive task than *Ss* low in threat-oriented affiliation motivation.

*Hypothesis II.* Under conditions that lead them to anticipate rejection, *Ss* high in threat-oriented affiliation motivation are less productive on a cooperative task than *Ss* low in threat-oriented affiliation motivation.

Under conditions that lead them to anticipate acceptance, *Ss* high in goal-oriented affiliation motivation should be more productive on a cooperative task and less productive on a competitive task than *Ss* low in goal-oriented affiliation motivation. The present experiment cannot adequately test these predictions concerning goal-oriented affiliation motivation, since the design did not include situations in which anticipation of acceptance was made salient. However, the effects of goal-oriented affiliation motivation may appear in a weaker form in the control groups which experienced no arousal of anticipated acceptance or rejection.

## METHOD

### *Subjects*

A total of 88 male undergraduates participated in the experiment. They were selected from several sections of an introductory psychology course according to their affiliation and achievement motivation scores, (*n* Affiliation, *n* Achievement) and met in groups of four, five, or six for the experiment. Individual groups were composed of *Ss* both high and low in *n* Affiliation.

<sup>1</sup> This report derives from a larger study submitted to the University of North Carolina as a doctoral dissertation. The author is indebted to Drs. John W. Thibaut, Edward E. Jones, and to the other members of the Organization Research Group for constructive criticism.

The study was conducted in the program of the Organization Research Group of the Institute for Research in Social Science, which is supported jointly by the Ford Foundation and the Office of Naval Research [Nonr-855 (04)].



The selection procedure assured comparable *n* Achievement scores for *Ss* in the various cells of the design.

### Design

In each experimental condition, the actual tasks to be performed by *Ss* were identical. Only the task instructions and the experimental procedures prior to the task were manipulated. Half of the groups received competitive task instructions; the other half, cooperative task instructions. The experimenter (*E*) attempted to arouse the affiliation motive and anticipation of rejection in half of the groups; for control, this procedure was omitted with the other half. Thus, the experimental procedures called for a two-by-two design, varying task instructions and arousal conditions. Within each of the four cells, *Ss* could be divided into those above and those below the median of the measures of motivation.

### Measurement of *n* Affiliation

During their first laboratory meeting the *Ss* were asked to write stories in response to picture stimuli according to the procedure described by McClelland *et al.* (5). A six-picture measure of *n* Affiliation was obtained by scoring according to the procedure described by Atkinson *et al.* (1). Five of the pictures have been used by McClelland and his associates. The sixth picture, representing four men presumably discussing something or chatting casually, was taken from Henry and Guetzkow (3).

Each story containing affiliation imagery was classified as either predominantly positive (goal-oriented) or predominantly negative (threat-oriented). A story was classified as positive when it contained predominantly approach responses, such as statements about positive affective relationships between characters, or characters striving for and attaining affiliation goals (positive outcome). Components considered positive were companionate activities, such as parties, reunions, visits, small talk, bull sessions, etc. A story was classified as negative when it contained predominantly avoidance responses, such as statements about concern over separation, or characters striving for an affiliation goal not specifically attained in the story (negative or questionable outcome). Negative components included concern over: being left out or rejected; being lonely or without friends; experiencing physical separation including death; experiencing psychic separation, such as disagreements or quarrels (including attempts to make amends for past actions that brought about separation or rejection); and experiencing no reciprocity of love.

The +Affiliation score of any *S* was the total of the *n* Affiliation scores for all stories which were judged to be predominantly positive. The -Affiliation score was derived in a corresponding manner. The sum of the positive and negative affiliation scores for an *S* (without regard to sign) thus equals the total *n* Affiliation score as described by Atkinson *et al.* (1).

The pictures were found to elicit approximately equal numbers of positive and negative stories and thus appear to yield a balanced measure of total *n* Affiliation. There was essentially no correlation between the positive and the negative affiliation motivation scores ( $r = -.15$ ,  $N = 88$ ).

Interscorer reliability of two experienced scorers<sup>2</sup> was measured throughout the scoring by spot checking approximately one-third of all stories scored. The agreement between scorers was 95 per cent for affiliation imagery alone, and 92 per cent for all categories.

### Affiliation Arousal

Each *S* who underwent this procedure was asked to rank a series of traits individually from most to least complimentary. An eight-minute group discussion followed in which the group was to try to agree on a single group ranking of the traits. Each *S* was told that he was also to form an opinion of the other members, since he would be asked to evaluate them and to choose work partners for a subsequent task. After the discussion, each *S* stood before the group and read his original ranking of the traits, while all others chose two characteristic traits for him. Finally, each *S* indicated his preference for work partners by ranking the others. This entire procedure was omitted in the control groups.

The arousal procedure followed closely that described by Atkinson *et al.* (1) with the addition of the group discussion. This was included in order to give the *Ss* some basis for rating each other and so that some of the characteristics of a group could begin to emerge. The ratings and partner choices were intended to make *S* acutely aware of being evaluated by other members of the group. Both Shipley and Veroff (6) and Atkinson *et al.* (1) point out that this brings out fear of rejection. Therefore, *Ss* in the affiliation arousal condition may be considered to anticipate rejection much more strongly than those in the control condition.

### Task Instructions

The competitive instructions presented an individual goal of competition for highest productivity, the results to be a ranking of individuals in the group. The cooperative instructions presented a team goal of cooperation among members on individual problems to be summated for the group score.

To make the cooperative task instructions more realistic, the group was instructed to divide the total task among themselves in a specified way so as to finish as many of the problems as possible, each person starting on a different page in the booklet, working the remaining pages in order, and returning to the first page if he reached the end. It was stressed that he should do as many problems as possible since this would increase the group score. The booklets were actually composed so that although each *S* started on a different page, they were all actually working on the same problems in the same order. The task was therefore comparable under all conditions.

### Group Attractiveness

On the post-experimental questionnaire, each *S* was asked to rate on a seven-point scale his agreement with items concerning his liking for the group during the experiment. The items were as follows: (a) If I had to choose a group to work with me on a real problem, I would choose some of the members of this group. (b) I

<sup>2</sup> Mrs. John S. Crittenden and the author.

TABLE 1  
MEAN NUMBER OF TASKS ATTEMPTED AND THREAT-  
ORIENTED AFFILIATION MOTIVATION

Task and Level of Threat- Oriented Affiliation Motivation	Mean Tasks Attempted	
	Affiliation Arousal (Threat of Rejection)	No Arousal
Competitive Task		
High -Affiliation	133.8	126.3
(n)	(10)	(8)
Low -Affiliation	112.5	115.0
(n)	(12)	(14)
Mean	122.2	119.1
(n)	(22)	(22)
Cooperative Task		
High -Affiliation	100.6	87.8
(n)	(11)	(10)
Low -Affiliation	110.2	84.6
(n)	(11)	(10)
Mean	105.4	86.3
(n)	(22)	(22)

felt some dislike for at least one of the members during the tasks. (Reverse scored.) (c) In general, I liked most of the people in my group very well.

### Productivity

The productivity score for each *S* consisted of the sum of the problems attempted on both tasks. Task I was a simple arithmetic task. Task II was a scrambled words task. Performance on both of these tasks has previously been found to be related to achievement motivation (4, 7).

### Procedure

The affiliation arousal was the first procedure for aroused groups and was omitted for control groups. The task instructions placing *S* in a competitive or a cooperative task situation were then read, after which *Ss*, seated in isolation booths, worked on 120 arithmetic problems for 12 minutes, followed by 12 minutes' work on 120 scrambled words. Feedback was supplied each *S* privately, informing him that his productivity ranking was just below the median rank of his group. This was done immediately prior to administration of the post-experimental questionnaire in order to equate effects of knowledge of results on the tasks. The questionnaire included the "Group Attractiveness Scale" and concluded the experiment except for an informal discussion with the *Ss*.

## RESULTS

Table 1 presents the productivity means for *Ss* high and low in -Affiliation under the four different experimental conditions. Considering only the experimental manipulations, analysis of variance demonstrates that mean productivity in the competitive task situation

TABLE 2  
MEAN NUMBER OF TASKS ATTEMPTED AND GOAL-  
ORIENTED AFFILIATION MOTIVATION

Task and Level of Goal- Oriented Affiliation Motivation	Mean Tasks Attempted	
	Affiliation Arousal (Threat of Rejection)	No Arousal
Competitive Task		
High +Affiliation	118.4	118.3
(n)	(9)	(14)
Low +Affiliation	124.8	120.5
(n)	(13)	(8)
Cooperative Task		
High +Affiliation	101.8	92.5
(n)	(12)	(13)
Low +Affiliation	109.7	77.4
(n)	(10)	(9)

was significantly higher than mean productivity in the cooperative task situation ( $p < .01$ ). There was no significant difference between mean productivity under affiliation arousal and under no arousal. However, considering only the cooperative task groups, mean productivity of *Ss* under affiliation arousal was significantly higher than productivity under no arousal ( $p < .01$ ).

A median split of +Affiliation or -Affiliation scores within each of the four experimental conditions resulted in unequal cell entries. A multiple regression analysis was therefore computed (see Table 3) to determine the effects due to level of motivation within each experimental condition. For descriptive purposes the means in Tables 1 and 2 for high and low levels of motivation will be discussed. However, significance levels will refer to tests of the variance due to regression.

Hypothesis I predicts higher productivity for *Ss* high in -Affiliation than for *Ss* low in -Affiliation on a competitive task under affiliation arousal (threat of rejection). Table 1 shows a large difference in the predicted direction, and regression analysis reveals that this is a reliable difference ( $p < .01$ ) confirming the hypothesis. Under no arousal there was a similar but insignificant difference.

The data in Table 1 are also consistent with Hypothesis II which predicts lower productivity for *Ss* high in -Affiliation than for *Ss* low in -Affiliation on a cooperative task under affiliation arousal. The regression analysis, however, reveals that this difference falls short of customary standards of significance ( $p < .07$ ).



TABLE 3  
ANALYSIS OF VARIANCE DUE TO REGRESSION OF PRODUCTIVITY ON GOAL-ORIENTED  
AND THREAT-ORIENTED AFFILIATION MOTIVATION SCORES

Arousal	Task	Motive	df	Mean Square	F	b
Affiliation Arousal	Competitive	+Affiliation	1	814	1.15	-2.1754
		-Affiliation	1	5,173	7.30***	+6.0783
	Cooperative	+Affiliation	1	547	.77	-1.7836
		-Affiliation	1	2,546	3.59**	-4.2643
No Arousal	Competitive	+Affiliation	1	110	.16	-0.8012
		-Affiliation	1	490	.69	+1.8714
	Cooperative	+Affiliation	1	2,027	2.86*	+3.4327
		-Affiliation	1	101	.14	+0.8500
	Error†		76	709		

\*  $p < .10$ .

\*\*  $p < .07$ .

\*\*\*  $p < .01$ .

† Twelve degrees of freedom were used in testing variance due to differences between cell means and variance due to regression.

For a test of the significance of the interaction of level of threat-oriented n Affiliation with kind of task under affiliation arousal, the difference between the two computed regression coefficients was tested. The significant relationship is found that would be predicted by combining Hypotheses I and II ( $p < .01$ ).

It has been noted that the experimental design did not provide a situation of heightened anticipation of acceptance for testing predictions for +Affiliation, but data for the No Arousal condition provide suggestive evidence. Table 2 presents means for Ss high and low in +Affiliation. As predicted, productivity on a cooperative task under No Arousal was higher for Ss high in +Affiliation than for Ss low in +Affiliation, although the trend falls short of statistical significance ( $p < .10$ ).

Analysis of productivity and total n Affiliation scores reveals smaller relationships similar to those with -Affiliation in the arousal cells, but +Affiliation and -Affiliation scores considered separately seem to be better predictors of performance than total n Affiliation under the present conditions.

#### DISCUSSION

The main hypotheses, which are in general supported by the data in Table 1, were based on the assumption that competitive productivity under certain conditions is the result of hostility induced in Ss high in -Affiliation by threat of rejection. Using the Group Attractiveness Scale as a rough measure of hostility toward the group, it was possible to examine the relationships between hostility,

-Affiliation, and the experimental manipulations. Although the relationships were all small and insignificant, they were remarkably consistent. Hostility was greater in Arousal than in No Arousal cells, and greater for Ss high in -Affiliation than for Ss low in -Affiliation. Competitive productivity followed the same pattern. It was higher in the Arousal cells, and (as shown in the test of Hypothesis I) high for Ss high in -Affiliation. Further, competitive productivity was higher for Ss high in hostility than for Ss low in hostility; and, conversely, cooperative productivity was lower for Ss high in hostility than for Ss low in hostility.

Table 1 also shows that cooperative productivity was significantly greater under Arousal than under No Arousal. This finding cannot be attributed to threat of rejection, which should inhibit cooperative productivity, especially in Ss most influenced by threat. It may well be attributable to the group discussion, in which only the aroused group participated and which allowed Ss to become acquainted. They seemed to gain a more positive attitude toward the situation, as indicated by post-experimental discussions with the groups, and this positive attitude could well have produced an increased incentive for cooperative work.

Further examination of the data shows that cooperative productivity under arousal was relatively greater for Ss low on any of the measures of affiliation motivation. The most striking difference was with total n Affiliation. Cooperative productivity for Ss low in n

Affiliation was higher in the aroused than in the control condition, (aroused mean = 116.8, control mean = 79.5), while productivity for Ss with high *n* Affiliation was approximately the same under the two conditions (aroused mean = 89.0, control mean = 90.5). The overall increase under arousal, which is attributable entirely to the Ss with low motivation, may be similar to Wendt's findings (7) with achievement motivation under scheduled and nonscheduled task conditions. He interprets his results as showing the effect of imposing extrinsic motivation upon Ss low in intrinsic motivation in the scheduled situation, thereby bringing their productivity up to those who were intrinsically motivated and had high productivity under either scheduled or nonscheduled conditions. In the present experiment, however, the productivity under arousal of Ss low in *n* Affiliation was significantly higher than that of the Ss high in *n* Affiliation. It would seem that the extrinsic threat of rejection has interacted with the intrinsic *n* Affiliation, thereby inhibiting increased cooperative productivity under the arousal condition for Ss high in *n* Affiliation, while the group discussion facilitated cooperative productivity for Ss low in *n* Affiliation to whom the threat was not as important.

#### SUMMARY

This study tested the hypotheses that under conditions favoring the anticipation of rejection, Ss high in threat-oriented affiliation motivation are (a) more productive on a competitive task, and (b) less productive on a cooperative task than Ss low in threat-oriented affiliation motivation. These hypotheses were in general supported by data from 88 undergraduate Ss.

In the control condition, which was treated neutrally with respect to arousal of anticipated acceptance or rejection, Ss high in goal-oriented affiliation motivation (as compared to

Ss low in this respect) tended to be more productive on a cooperative task. This relationship did not reach statistical significance.

Under Affiliation Arousal (which included both a group discussion and threat of rejection) there was higher cooperative productivity than under No Arousal, a difference contributed almost entirely by Ss low in total affiliation motivation (goal-oriented plus threat-oriented). This result was interpreted as showing an interaction between the effects of the group discussion and threat of rejection. The group discussion facilitated cooperative productivity in general, but among Ss high in affiliation motivation this effect was counteracted by the threat of rejection.

The other relationships between overall affiliation motivation and productivity were insignificant. The importance of distinguishing between goal- and threat-oriented affiliation motivation was stressed.

#### REFERENCES

1. ATKINSON, J. W., HEYNES, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405-410.
2. FRENCH, ELIZABETH, G., & CHADWICK, IRENE. Some characteristics of affiliation motivation. *J. abnorm. soc. Psychol.*, 1956, **52**, 296-300.
3. HENRY, W. E., & GUETZKOW, H. Group projection sketches for the study of small groups. *J. soc. Psychol.*, 1951, **33**, 77-102.
4. LOWELL, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1952, **33**, 31-40.
5. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
6. SHIPLEY, T. E., JR., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349-356.
7. WENDT, H. W. Motivation, effort, and performance. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. Pp. 448-459.

Received October 25, 1956.



# COGNITIVE SET AND MOTIVATIONAL FACTORS IN THE PERCEPTION OF NEUTRAL AND THREAT-RELATED STIMULI<sup>1</sup>

RUTH C. WYLIE

*Sarah Lawrence College*

RECENTLY, experimental psychologists have been interested in the proposition that perceptual thresholds are higher for so-called emotionally threatening stimuli, and lower for positively valued stimuli than for affectively neutral ones. Tests of such a proposition have yielded ambiguous results because of difficulties in controlling or knowing the status of such factors as: (a) subjects' attitudes toward the stimuli, (b) frequency of subjects' past experiences with the stimuli (c) subjects' cognitive sets or expectancies; and (d) the availability of the required responses as they are determined by subjects' previous practice in making the responses and willingness to report honestly (1-3; 5-11). This experiment employed new means to control the above-mentioned variables in a comparison of the influence of motivational factors plus cognitive set with the influence of cognitive set alone. In this design the experimenter (*E*) induced the desired emotional state and cognitive expectancies in the subjects under laboratory control; and the use of pretested numbers as stimuli, presented under specially devised conditions, provided improved control of factors *b*, *c*, and *d*.

The following hypotheses were tested:

*Hypothesis 1.* When subjects (*Ss*) have just been given an auditory presentation of a stimulus and have reason to expect that they may be given the same stimulus visually, their duration thresholds for such visual stimuli are lower than for otherwise comparable visual stimuli for which no such cognitive set has been established.

*Hypothesis 2.* When *Ss* desire to succeed in a task, their duration thresholds for visual stimuli indicating that they have failed are greater than their duration thresholds for otherwise comparable visual stimuli indicating

that they have succeeded. These differences are expected to be greater than could be accounted for on the basis of cognitive set or expectancy alone.

*Hypothesis 3.* When *Ss* desire to succeed in a task, the content of their prerecognition guesses about visual stimuli which indicate that they have failed resembles the content of stimuli which, had they been presented to *S*, would have indicated success. This distortion of prerecognition guesses is expected to be greater than could be accounted for on the basis of cognitive set or expectancy alone.

## METHOD

### *Stimulus Materials*

Two-place numbers were used as stimuli on the assumptions that: (a) such stimuli would be emotionally neutral to *Ss* until experimental conditions produced an emotional set toward them; (b) the response availability for all two-place numbers would be the same, insofar as this is determined by frequency of past visual exposure, practice in pronunciation, and lack of embarrassment in pronunciation. On the basis of a preliminary experiment, testing the tachistoscopic recognizability of two-place numbers of a standard print type and size, these numbers were classified into three groups: hard, medium, and easy to recognize, for use in matching the "threatening" and "rewarding" numbers in the main experiment.

### *Subjects*

The *Ss* were 51 volunteer female college students, naive with respect to the purpose of the experiment. They were divided into three groups of 17 each.

### *Procedure*

*Group 1.* Group I *Ss* were told they were to participate in two separate experiments. In the "first experiment" (Task 1), each *S* was given individually a specially devised and pretested task. She was told that, although the parts of the task were simple, this was a new intelligence test which revealed the ability of a person to keep her head and concentrate in a fast-moving, distracting situation which might "rattle" her. She was instructed that preliminary information indicated that this is a very important kind of intelligence for everyday living, and she was encouraged to keep trying by being told she would receive partial credit for nearly correct answers. This task consisted of 14 chains of arithmetic operations, with ten operations per chain. The *E* read each chain aloud at the rate of about one operation every five seconds, while *S* did the

<sup>1</sup> The Gerbrands tachistoscope used in this study was purchased under a grant from the American Philosophical Society. Thanks are due also to Dr. Georgiana Wylie who pretested the ego-involving task at another college.

operations mentally and wrote only her final answer for each chain. The oral presentation and mental computation prevented *S* from knowing whether any answer was right or wrong unless she had completely failed to give an answer or had obtained a number with more than two places. The *S* was asked to guess how many problems she had done correctly and how many almost correctly; and to tell how sure she was of this guess. (No *S* was sure of her guess.) Each *S* was asked to rate herself on her ability on this type of task, as compared with the average college student. Finally, she was given fictitious norms which were higher than the actual attainment of the college group. Interviews at the end of the entire experimental procedure indicated that *Ss* were ego-involved on Task 1.

After Task 1, *S* was told that the "first experiment" was finished and she was given a brief rest period before the "second experiment." Meanwhile *E* prepared, in another room, a series of 14 numbers, to be presented in a Gerbrands mirror tachistoscope. Regardless of *S's* answers on Task 1, half the numbers for this series were selected to agree and half to disagree with *S's* answers. The disagreeing numbers were chosen to be at least as readable as the agreeing ones. Agreeing and disagreeing items were put in random sequence to control for practice and fatigue.

The *S* was then brought into the tachistoscope room and asked if, while she was in the laboratory anyway, she would mind helping *E* test a new apparatus. An effort was made to avoid ego-involving *S* in being very accurate in the perceptual task as such, and to keep her from suspecting the connection between Tasks 1 and 2 by telling her that *E* was only testing the lighting conditions and apparatus for use in a further experiment, and was merely trying to make this perceptual task less boring to *S* by using for stimuli the correct answers to Task 1. (The purpose of avoiding ego-involvement in perceptual accuracy was to prevent masking possible motivationally-induced distortions, something which might happen if the accuracy set for the perceptual task were too high.) As part of the detailed instructions, *S* was told to report what she saw on the screen after each successive exposure, basing her guesses on the visual characteristics of the stimuli, *not* on any other method of guessing. The *S* was given a 45-sec. practice period in reading aloud the numbers 0 through 9, in various scrambled orders. She never removed her eyes from the eyepiece during the instructions and practice period, thus achieving adaptation to the level of illumination used during the experiment.

Before the exposure of the first stimulus number from the experimental list, *E* read aloud *S's* answer to Problem 1 on Task 1 and told her that if the first tachistoscope number was the same as her answer, this would mean that her answer to Problem 1 was correct. Beginning at .01 second, *S* was given successively longer exposures of the number until she made two successive veridical reports, the first of which was taken as the threshold. The duration thresholds in hundredths of a second and all prerecognition guesses were recorded by *E*. This procedure was repeated for each of the remaining 13 numbers.

At the end of Task 2, the entire experiment was explained to *S* and she was interviewed to see if she had become ego-involved and to find out whether she had

discerned the true connection between Tasks 1 and 2. The *Ss* reported that the directions had been convincingly misleading, but some of them said they had wondered whether hearing the number and/or wishing to see it might have influenced their perceptions. However, they all reported trying hard to base their guesses exclusively on the visual appearance of the stimuli. Each *S* was asked to keep the experiment secret so that future *Ss* would be naive.

It can be seen from the description of the procedure used with Group I that both expectancy and "wishful thinking" might influence their perceptual thresholds for those tachistoscope numbers which agreed with the numbers *E* read out to them.

*Group II.* Each Group II *S* was arbitrarily paired with a Group I *S*. The treatment given to Group II *Ss* was intended to make their experiences in the tachistoscope room exactly the same as those of Group I, except that their attitudes toward the numbers used as visual stimuli should be emotionally neutral, allowing one to study the influence of cognitive expectancies in the absence of motivational influences. Consequently, these *Ss* heard nothing about Task 1, but went directly to the tachistoscope room where they were given the same instructions for Task 2 as had been given to Group I, with the following exception: Group II *Ss* were told that *E* would "read out a number" instead of that *E* would "read out your answer to a problem on the intelligence test you just took." Any given Group II *S* was given the same list of "read out" and "viewed" numbers as had been used for her corresponding Group I *S*. Each Group II *S* was told, "I can't tell you how many of the numbers flashed on the screen will be the same as those on my list and how many will be different. There might be any degree of agreement between zero and 100 per cent. The greatest likelihood is that there will be about — agreements out of 14 numbers which you will have flashed on the screen. You can't be sure of this, however." The number of agreements suggested to a given Group II *S*, was the number of answers the corresponding Group I *S* stated she thought she had gotten correct at the end of her Task 1 test.

A careful inspection of the procedures given Groups I and II suggested that any differences in performance between these groups were to be interpreted with caution, because the parallel instructions to the groups did not suffice completely to equate cognitive expectancies between them. The *Ss* in Group I, for example, had reason from Task 1 to *expect* (as well as to *hope*) that some of the numbers might be *nearly* agreeing, if not exactly agreeing, even though the tachistoscope directions did not mention this possibility. On the other hand, the Group II *Ss* had had no comparable previous experience that might have established a cognitive expectancy of a certain number of *near* agreements. In an attempt to equate this factor of cognitive expectancy more effectively, the procedure used for Group II was administered to a third group, with the single modification that the directions were changed by the addition of the underlined clause below: ". . . The greatest likelihood is that there will be about — agreements out of 14 numbers which you will have flashed on the screen, with perhaps — coming fairly close." The exact number mentioned in the underlined clause was that which the corresponding Group I *S*



thought she had gotten nearly correct in Task 1. By this change of directions the expectancy of number of *near* as well as exact agreements was made more comparable between each Group III *S* and her matched Group I counterpart.

### RESULTS

For each *S*, the median threshold for all agreeing numbers was found (called *A*) and the median threshold for disagreeing numbers was obtained (called *D*).<sup>2</sup> (*A* - *D*) was predicted to be negative in the cases of all three groups. Groups II and III were employed to test Hypothesis 1 which was confirmed by the finding that 31 out of 34 *Ss* in these groups showed negative values for (*A* - *D*), a proportion significant far beyond the .01 level. Since other variables have been controlled, this result can reasonably be attributed to the effects of hearing the number and expecting the possibility of seeing it, i.e., to cognitive sets.

Although simply hearing the number and expecting possible agreement leads to faster perception, the second hypothesis asserts that *hoping* to see what one has just heard and *hoping not* to see any other number should make the (*A* - *D*) values of the Group I *Ss* more negative than those of Group II or Group III *Ss*. In order to make such comparisons, a refined measure of the discrepancy between thresholds for agreeing and disagreeing items was needed that takes into account the individual differences in intra-individual variability in thresholds, in terms of which the groups were not equated. To obtain such a measure, an index of intra-individual variability (here called *I*) was first computed for each *S* by taking the range of her middle six thresholds.<sup>3</sup> Her (*A* - *D*) value could then be evaluated in relation to her overall intra-individual variability. The score used

as the dependent variable in testing Hypothesis 2 may be summarized thus:

$$\frac{(A - D)}{I}$$

where

*A* = *S*'s median threshold for agreeing numbers

*D* = *S*'s median threshold for disagreeing numbers

*I* = range of middle six thresholds for *S*

The difference in mean score between Group I and Group II was found to be +0.11, which yielded an insignificant *t* value of 0.60<sup>4</sup>, while the difference in mean score between Group III and Group I was only -0.01. Thus, the data give no support to Hypothesis 2; i.e., there is no indication that wishful thinking added to expectancy produces any greater effect on duration thresholds than does expectancy alone.

The third hypothesis postulated that pre-recognition guesses of the Group I *Ss* might reveal "wishful thinking" by tending to be closer to the "read-out" number than were the pre-recognition guesses of Group II or Group III *Ss*. To test this hypothesis, all pre-recognition guesses within  $\pm 10$  of the "read-out" number (*S*'s answer in the case of Group I), were tallied and compared to the total number of pre-recognition guesses made by *S*, yielding a proportion score for each *S*. The differences between the proportion score of each Group I *S* and that of her corresponding Group II *S* were obtained by subtraction across pairs. The mean difference was 0.14 and *t* = 2.48, which is significant between the .02 and .01 level on a one-tailed test. These results support the hypothesis that the Group I *Ss*, who are hoping to see a number at least *near* their answer, will tend to see more numbers near the "read-out" number than will the control *Ss*, who have no reason to care whether the tachistoscope number is near the "read-out" number. However, for reasons described under Method, a comparison of Group III with Group I provides a better controlled test of Hypothesis 3. When the pre-recognition guesses of Group III were compared to those of Group I in the same way

<sup>2</sup> An item on which a Group I *S* knew she had failed because she got no answer or got more than a two place number for an answer was eliminated from the calculation of *D* of both the Group I *S* and the corresponding Group II *S*. To keep the counterbalancing of the ordinal position of agreeing and disagreeing items intact, in such a case, the nearest agreeing item was also eliminated from the calculations of *A* for the Group I and corresponding Group II *S*.

<sup>3</sup> This measure of variability was used rather than the *SD* because: (a) the number of thresholds in each individual's distribution was small, (b) the individual distributions were often skewed. (*I* approximates the interquartile range without involving interpolations or decimal values unwarranted by the size of the distributions or precision of the data.)

<sup>4</sup> No correlation on the dependent variable was found between Group I and Group III *Ss*, so uncorrelated *t* formulae were used.

that II and I had been compared, the difference was  $+ .08$ , which favored the hypothesis at the  $.10$  level (one-tailed test). Groups II and III did not differ significantly between themselves as to prerecognition guesses, so they were combined and then compared with Group I. The difference was  $+ .11$ , with a  $P$  slightly  $> .05$  on a one-tailed test.

#### DISCUSSION

The results of this experiment give clear-cut evidence for the effects of cognitive expectancies upon duration thresholds, whether or not motivational influences which might operate in the same direction were present. The data give some support to the notion that prerecognition guesses may be modified by motivational influences such as desire to succeed in an ego-involving task. No support is given the proposition that such motivational influences affect duration thresholds when other variables are held constant (the usual definition of perceptual defense).

The question arises as to whether those differences which were found are attributable to effects on the perceptual processes or to effects on the response processes, to follow the distinction developed most recently by Garner, Hake, and Ericksen (4). It seems likely that the effects are on the perceptual processes in view of the operations instituted to control for the influences on response processes. In the first place, the use of numbers should make motivational reasons for withholding responses less than they would be in the case of taboo words. In addition, the choice of two-place numbers provided stimuli which the Ss had had much practice in pronouncing. And finally, the instructions to base guesses solely on the visual properties of the flash should tend to reduce guessing behavior from other sources.

The results on cognitive set are in line with much previous work, and they further point up the need for controlling for this factor carefully in all studies where it is not the independent variable.

The data on the role of motivation are of interest in the light of the current controversy concerning the status of perceptual defense and accentuation. A number of investigators have suggested that the positive experimental findings in this area are attributable to uncontrolled variables such as set, frequency of

S's past experience with the stimuli and the like. The effects obtained are sometimes diminished or even reversed when these other variables are altered (6, 10, 11). In one well controlled recent experiment, no evidence of perceptual defense could be obtained (13). It was hoped that the new methods and stimulus materials introduced in this study would also effectively control the variables which have led to ambiguity in previous studies. Insofar as this aim has been realized, these results give no evidence of motivational effects on duration thresholds under comparable conditions of cognitive expectancy.

The present findings, and those of other experiments, may perhaps be reconciled with the concept of perceptual defense on the assumption that perceptual defense occurs only within certain limiting conditions. So far, these conditions remain to be explored and specified precisely, and one value of this experiment lies in suggesting what some of them may be. First, it is possible that laboratory-manipulated ego-involvement, while desirable for precision of control, may produce motivation too weak to affect duration thresholds. Since the motivation was apparently strong enough to exert some effects on the prerecognition guesses, one might infer that different minimum strengths and/or durations of motivation may be required to affect different measures of perceptual distortion.

It might be argued, second, that motivational effects do not appear when some other, more influential variable such as cognitive set is present to override them. On this view, the present results could be interpreted as analogous to the findings of Postman and Schneider (12) on familiarity, from which they concluded that differences in value rank produced significant differences in perceptual sensitivity to unfamiliar but not to highly familiar words.

Finally, there is the question whether the findings on perceptual defense may vary as a function of types of material used. Perhaps affect can become attached to some stimuli (e.g., words and pictures) more easily than to numbers. If so, the psychological variables leading to such differences among materials need to be determined.

In any event, the present findings suggest that caution and further empirical qualifications are in order before making generalizations about the occurrence of perceptual



defense, and about the variables which limit it or with which it interacts, if and when it occurs.

#### SUMMARY AND CONCLUSIONS

An experiment testing the following three hypotheses was described:

1. When Ss have just been given an auditory presentation of a stimulus and have reason to expect that they may be given the same stimulus visually, their duration thresholds for such visual stimuli are lower than for otherwise comparable visual stimuli for which no such cognitive set has been established.

2. When Ss desire to succeed in a task, their duration thresholds for visual stimuli indicating that they have failed are greater than their duration thresholds for otherwise comparable visual stimuli indicating that they have succeeded. These differences are expected to be greater than could be accounted for on the basis of cognitive set or expectancy alone.

3. When Ss desire to succeed in a task, the content of their prerecognition guesses about visual stimuli which indicate that they have failed resembles the content of stimuli which, had they been presented to S, would have indicated success. This distortion of prerecognition guesses is expected to be greater than could be accounted for on the basis of cognitive set or expectancy alone.

The design attempted to control by experimental manipulation variables which have caused difficulty in previous experiments in this field, i.e., S's attitude toward the stimulus materials, frequency of his past experience with perceiving the stimulus materials and pronouncing them, inherent readability of the stimulus forms and willingness to report honestly.

The influence of cognitive set on duration thresholds was demonstrated (Hypothesis 1).

No evidence was found for the effects of motivation on duration thresholds (Hypothesis 2), but some evidence was found for the influence of motivation on the content of prerecognition guesses (Hypothesis 3).

#### REFERENCES

1. COWEN, E. L., & BEIER, E. G. Threat, word frequencies, and recognition. *J. abnorm. soc. Psychol.*, 1954, **49**, 178-182.
2. ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175-182.
3. FREEMAN, J. T. Set or perceptual defense? *J. exp. Psychol.*, 1954, **48**, 283-288.
4. GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149-159.
5. HOWES, D. On the interpretation of word frequency as a variable affecting speed of recognition. *J. exp. Psychol.*, 1954, **48**, 106-112.
6. LACY, O. W., LEWINGER, NATALIE, & ADAMSON, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169-174.
7. LAZARUS, R. Is there a mechanism of perceptual defense? A reply to Postman, Bronson, and Gropper. *J. abnorm. soc. Psychol.*, 1954, **49**, 396-398.
8. NEISSER, U. An experimental distinction between perceptual process and verbal response. *J. exp. Psychol.*, 1954, **47**, 399-402.
9. POSTMAN, L. On the problem of perceptual defense. *Psychol. Rev.*, 1953, **60**, 298-306.
10. POSTMAN, L. The experimental analysis of motivational factors in perception. In M. R. Jones (Ed.), *Current theory and research in motivation: A symposium*. Lincoln, Nebr.: Univer. of Nebraska Press, 1953. Pp. 59-107.
11. POSTMAN, L., BRONSON, W. C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215-224.
12. POSTMAN, L., & SCHNEIDER, B. H. Personal values, visual recognition and recall. *Psychol. Rev.*, 1951, **58**, 271-284.
13. TURKEL, H. Perceptual defense in word recognition: A test of its generalized occurrence. Paper read at Eastern Psychol. Ass. March, 1956.

Received November 12, 1956.

# MATERNAL CHILD-REARING PRACTICES AND AUTHORITARIAN IDEOLOGY<sup>1</sup>

I. HART<sup>2</sup>

Duke University

THIS study was an attempt to investigate maternal preferences in techniques for controlling child behavior as a function of socialization situations and a particular aspect of maternal personality. The basic hypothesis was that patterns of child-rearing preferences can be predicted from personality trends expressed in ideological form; specifically, authoritarianism. As a test of this hypothesis, specific predictions were formulated, derived from the clinical insights and descriptive statements of Adorno *et al.* (1) and Levinson and Huffman (4). These predictions were tested in terms of the socialization model of Whiting and Child (7).

Writers on the authoritarian personality have contended that certain child-rearing practices are both antecedents and consequents of authoritarianism. The authoritarian parent is said to emphasize discipline that takes the form of "bodily harm, social isolation and/or shaming, rather than the loss of love, since the giving of love and the establishment of strong affectional ties are not likely to be primary issues in an autocratic setting" (4).

Whiting and Child's analysis of the punishment techniques customarily used by parents in socializing their children bears considerable congruence to the range of behaviors that Levinson and Huffman attribute to high and low authoritarian parents. One phase of Whiting and Child's analysis rests on the assumption that punishments differ in the extent to which they contribute to maintaining the child's orientation to the goal of parental affection. Thus they distinguish three kinds of discipline as "love-oriented," i.e., serving to maintain the child's striving for parental love: denial of love, threats of denial of reward, and threats of ostracism. In contrast to these

techniques, they propose three kinds of discipline as having the overall effect of establishing tendencies in the child to avoid the parents. These "nonlove-oriented" punishment techniques are: physical punishment, threats of physical punishment, and punishment by ridicule. Whiting and Child did not include actual denial of rewards and actual ostracism in either the love-oriented or the nonlove-oriented categories because these techniques were considered to contribute potentially both to striving for parental affection and to the development of avoidance tendencies. These latter punishments were described as "ambiguous." On the basis of the descriptions of Adorno *et al.* and of Levinson and Huffman, low and high authoritarian parents might be expected to select love-oriented and nonlove-oriented disciplinary techniques, respectively.

## METHOD

One hundred and twenty-six mothers, who had children between the ages of 2½ and 5½, were studied by means of intensive but structured interviews in which each mother was asked to indicate her most probable response to her own child's behavior in specific situations. Free responses were submitted to three independent raters (psychology graduate students) for coding appropriate to the Whiting and Child categories of disciplinary techniques, according to agreement of two out of the three raters. The subjects (Ss) were white, native born, Christian Americans ranging in age from 23 to 43. Mean mother's age was 30.8; mean age of the children was 4.1.

In an attempt to keep the effects of education and socioeconomic level as constant as possible, the sample was restricted to middle-class mothers who had completed at least one year of college. The fathers were predominantly college graduates, and for the most part were engaged in either professional or higher level managerial occupations. The Ss were recruited by asking interviewees to suggest mothers who could participate in the study. Almost all of the Ss were able to suggest others; 95 per cent of those who were asked participated.

The interview schedule contained 38 items depicting various child behavior situations, and a modified version of Whiting and Child's categorization of punishment techniques, combined in questionnaire form. The behavior situations were chosen on the basis of a priori judgment and some brief preliminary research, in the expectation that they would occur rather frequently and tend to elicit some form of discipline rather than

<sup>1</sup> This article is based on a thesis submitted to Duke University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer expresses his appreciation to Dr. E. E. Jones for his helpful criticisms in the design and execution of the study.

<sup>2</sup> Now at Cowell Memorial Hospital, University of California, Berkeley.



reward from parents of children in the age range sampled. It was further intended that each situation represent an instance of behavior in one of the following six areas:

1. Feeding and oral activities: e.g., "insists on eating with fingers," "seldom finishes meal," etc.

2. Cleanliness-toilet training: e.g., "frequently has toilet accidents because won't interrupt play."

3. Sex: sexual curiosity, sexual and obscene remarks, handling genitals, body exposure.

4. Aggression: Items involve insolence and impudence, hitting and pushing, temper tantrums, explicit disobedience, willful property damage.

5. Dependence: These items attempt to depict behavior suggesting that the child lacks self-reliance, is attempting to gain parental support, etc., e.g., "whines or begs for help unnecessarily."

6. Independence: This area includes situations that are intended to suggest that the child is showing initiative, curiosity, assertiveness, e.g., "insists on own way with other children," "wanders away from home," etc.

In addition to the structured interview phase, each mother was asked to describe her three biggest problems concerning her child, to answer two open-end projective-type questions, and to fill out a combined version of the short-form (TFI) scale (4) and F scale (1).<sup>3</sup> The scores thus obtained (M, 3.70; SD, 1.09) provided the measure of authoritarian personality. The interviews took from 1½ to 2½ hours.

## RESULTS

Authoritarianism was found to be unrelated to each of the following variables: age of the mother, age of the child, sex and birth order of the child, and number of children in the family. The behavior situations included in the interview schedule were quite comparable to and representative of the kinds of behavior problems reported by the Ss.

The main results are described below with reference to three experimental predictions.

1. *The selection of nonlove-oriented disciplinary techniques should be a positive function of maternal authoritarianism score.* To test this prediction, a product-moment correlation was computed between each mother's authoritarianism score and the number of situations for which she selected nonlove-oriented techniques. The correlation thus obtained was .63 ( $p < .001$ ).

For a more detailed analysis, the distribution of authoritarianism scores was divided into thirds, with cutoff scores for highs and

<sup>3</sup> An abbreviated F scale of 12 items was used; it contained Items 4, 9, 13, 18, 21, 25, 26, 34, 37, 42 from Form 45, Item 44 from Form 60, and Item 55 from Form 78 of the original F scale (1). The TFI items were the 12 comprising the short form presented by Levinson and Huffman (4).

TABLE 1  
MEAN PERCENTAGE OF MOTHERS AT EACH LEVEL OF  
AUTHORITARIANISM WHO SELECT EACH  
TYPE OF DISCIPLINE

Type of Discipline and Level of Authoritarianism	Behavior Area						Overall Mean
	Independence	Dependence	Cleanliness	Feeding	Sex	Aggression	
Love							
High	30.33	65.12	59.52	47.29	70.26	18.64	48.53
Medium	27.90	69.52	69.64	58.93	80.57	23.14	54.95
Low	45.05	84.76	88.69	83.31	89.12	42.52	72.24
Mean	34.43	73.13	72.62	63.18	79.98	28.10	
Nonlove							
High	60.55	23.10	33.93	30.57	20.62	69.40	39.69
Medium	53.02	16.45	26.19	20.48	11.17	47.57	29.15
Low	47.60	7.36	8.93	8.81	8.14	41.88	20.45
Mean	53.72	15.64	23.02	19.95	13.31	52.95	
"Ambiguous"							
High	9.12	11.78	6.55	22.14	9.12	11.96	11.78
Medium	19.08	14.03	4.17	20.59	8.26	29.29	15.90
Low	7.35	7.88	2.38	7.88	2.74	15.60	7.31
Mean	11.85	11.23	4.36	16.87	6.71	18.95	

lows falling at 4.1 and 3.2, respectively. The relationship between mothers' authoritarianism and preferred disciplinary techniques may be observed in the right-hand marginal column of Table 1. The body of Table 1 lists the percentages of type of discipline selected by mothers of a given authoritarianism level, in the various behavioral areas. Mothers at all levels of authoritarianism in general selected more love-oriented than nonlove-oriented control measures. However, it will also be seen that as authoritarianism level increases, nonlove-oriented techniques are preferred increasingly, and that as authoritarianism level decreases, the tendency to select love-oriented techniques increases. Since practically all of the previous research concerning authoritarianism has employed only the extremes of the distribution, the confirmation of this prediction also lends support to the contention that authoritarianism is an intrinsic continuum, not merely a dichotomous or trichotomous aggregate of traits.

2. *High authoritarians in comparison with low authoritarians, should select relatively more nonlove-oriented responses for the sex and aggression areas respectively, than for the remaining areas.* This prediction was based on the supposition of authoritarian personality

theory that children's sexual and aggressive behavior are especially unacceptable to authoritarian parents, and especially subject to control by discipline.

If nonlove-oriented measures are the characteristic disciplinary techniques of high authoritarian parents, and if children's sexual and aggressive behavior is especially unacceptable to them, then it might be expected that authoritarian parents would select more nonlove-oriented responses for these areas than for the other areas. To test that portion of the prediction pertaining to aggression, each high and low authoritarian mother was given a score based on the differences between the percentages of nonlove-oriented responses in the aggression area and the combined value of the remaining areas (minus sex). The *t* ratio of 1.8 between the mean scores of the highs and lows, was close to significance at the .05 level.

The failure of the data to attain an acceptable level of significance may reflect the arbitrary grouping of genotypically different items in the particular categories adopted. The behavior situations classed under aggression included items differing markedly from each other in interpersonal significance. Perhaps behaviors embodying threat to authority relationships may be more germane to the authoritarian personality dimension than the particular situations chosen for inquiry. For example, one group of aggressive behaviors that may be interpreted as predominantly involving threat to parental authority discriminated significantly between highs and lows in an individual situation analysis, while the remainder did not. Aggression (and other areas of problem behavior as well) may require finer analysis to yield satisfactory categories in terms of interpersonal significance.

Not only did the data fail to support that part of the prediction pertaining to sex, but as Table 1 indicates, the differences between highs and lows in the selection of nonlove responses were smaller in the sex area than in any other area. It might be speculated that this finding indicates considerable enlightenment in the attitudes of the sample studied towards sexual behavior in their children. The writer was impressed with the casualness and acceptance which the mothers displayed concerning sexual manifestations in their children. Many of the mothers indicated that their own

mothers would have been shocked by such behavior, and would not have been so tolerant. While this suggests a cultural trend toward increased leniency in middle-class attitudes toward the expression of sex, it also is very likely in considerable part a reflection of the cultural sophistication of the sample. Moreover, sexuality per se was generally not recognized nor accepted as intrinsic in the particular behavior, but the behavior was generally equated to intellectual curiosity, or otherwise rationalized as part of the childhood learning process.

As may be observed in Table 1, the class of behavior control techniques preferred varied with behavior area. To determine if these variations exceeded chance probability, analyses of variance were computed in which the six behavior areas and the three authoritarianism levels were the independent variables and the classes of disciplinary techniques were the dependent variables. The data summarizing the analysis of variance for nonlove-oriented disciplinary techniques are presented in Table 2. Since love and nonlove response frequencies are essentially mirror images of each other, the results of the analysis of variance for love-oriented discipline were substantially the same. Since the distributions were heterogeneous in variance and form, the *F* ratios were interpreted as inflated. While the very large *F* ratio (137) suggests that it would be quite safe to assume that the main effect of behavior areas is significant, the comparatively small *F* ratio (2.2) for interaction effects does not warrant the conclusion of a significant interaction. This finding corresponds to the nonsignificant results for Prediction 2.

Inspection of Table 1 indicates that nonlove-oriented responses were selected most frequently in the independence and aggression areas, and least frequently in the sex and dependence areas. One possible explanation is that the independence and aggression areas depict behaviors that are more threatening to parental control and authority. However, many of the mothers explained their choice of response to the independence items referring to certain behaviors as potentially dangerous to the child by indicating that threat to their child's safety was the one time they did not hesitate to resort to physical punishment. It might be surmised that ag-



TABLE 2  
ANALYSIS OF VARIANCE OF PREFERENCE FOR NONLOVE-  
ORIENTED DISCIPLINARY TECHNIQUES

Source	df	SS	MS	F	P
Levels of authoritarianism	2	46796.62	23398.31	33.52	.01
Subjects within levels	123	85864.80	698.09		
Behavior areas	5	217209.28	43441.86	137.07	.01
Level $\times$ Area	10	7014.05	701.41	2.21	.05
Area $\times$ Subjects within levels	615	194917.34	316.94		
Total	755	551802.09			

gressive behavior on the part of the child might provoke parental aggressive retaliation. The most frequent reason advanced by the mothers for their reluctance to select nonlove-oriented responses for dependency behavior was fear that these techniques might cause the child to feel rejected.

3. *The low authoritarian's description of her child-rearing philosophy and related attitudes should contain more attitude statements with the judged consequence of maintaining the child's approach tendency to her, than should the high authoritarian's.* The high authoritarian's "philosophy" is expected to be judged as more likely than the low authoritarian's to produce and strengthen an avoidance tendency in the child. For the purpose of testing this prediction, each mother was asked the following open-end questions: "What is your philosophy of child-rearing?" and "What do you think are the most important things to take into consideration when dealing with child behavior problems?" To reduce the scope of the task of rating, only the protocols of the mothers with the 30 highest and lowest authoritarianism scores were used in the statistical test. These protocols were submitted to two additional judges who rated them in terms of the anticipated effect the depicted behavior would have on the tendency of the child to approach or avoid the mother. The protocols were in no way identified with regard to authoritarianism and, in fact, the judges were unacquainted with authoritarian theory or the hypotheses of the study. Each protocol was divided into meaningful, coherent idea units on a purely a priori basis, in an attempt to facilitate comparability of ratings. The task of the judges was to sort each unit of the mother's statements into an "approach" category, i.e., behavior thought to strengthen the child's

approach tendency; an "avoidance" category; or a "miscellaneous" category. Although the number of units in the protocols varied, the mean number of units of the highs was close to that of the lows. However, as a precaution and in an attempt to compensate for unequal length of protocols, the score for each protocol consisted of the summed approach rating minus the summed avoidance rating. Thus, the final score for each protocol is a discrepancy score and is assumed to represent the relative excess of approach over avoidance (or vice versa).

Fairly high interrater agreement is reflected in the product-moment correlation of .72 for the lows and .73 for the highs. It was concluded that there was sufficient reliability in the judges' ratings to warrant their combination in mean judgments. The mean discrepancy score of the lows was 36.1, *SD*, 7.6; that of the highs 30.5, *SD*, 6.42; each case representing an excess of approach over avoidance judgments. The *t* test computed to determine the significance of the difference between the scores of high and low authoritarians yielded a value of 3.08,  $P < .01$ , in support of Prediction 3.

The confirmation of this prediction suggests that the structured interview schedule has construct validity (2), and that the socialization techniques differentially preferred by high and low authoritarians have conceptual significance in terms of approach and avoidance principles.

An attempt was also made to determine if maternal child-rearing preferences varied significantly as a function of such variables as sex and age of the child, and if interaction between either of these variables and authoritarianism could account for differences in the preferences. The child's age range was arbitrarily divided into three groups:  $2\frac{1}{2}$ - $3\frac{1}{2}$ ;  $3\frac{1}{2}$ - $4\frac{1}{2}$ ;  $4\frac{1}{2}$ - $5\frac{1}{2}$ , and three analyses of variance were carried out, in which the three age ranges and three levels of authoritarianism were the independent variables, and the three disciplinary techniques were successively the dependent variables. Disciplinary techniques varied significantly as a function of age only in the case of "ambiguous" responses. There was a slight, but significant, tendency for "ambiguous" responses to be selected increasingly by all three groups, as child's age increased (*F* ratio 3.9,  $P < .05$ ).

In the case of nonlove-oriented discipline, but only then, child-rearing preferences varied significantly as a function of child's sex. All three authoritarian groups responded slightly more often with nonlove-oriented responses when the children were boys, than when they were girls ( $F$  ratio 5.9,  $P < .05$ ). As in the case of the age variable, there was no significant interaction between child's sex and authoritarianism.

#### DISCUSSION

These findings emphasize a conclusion that many have long taken for granted. They show that the choice of behavior control techniques, in a variety of socialization situations, is to a significant extent conditioned by measurable personality factors in the mother. Whiting and Child (7) have argued that parental personality is one critical link in the mediation of culture from generation to generation. The present study, like theirs, was not designed to clarify the antecedents and consequents involved in this process. However, the results do suggest that different emphases or alternative customs of the culture are mediated as a function of maternal personality.

The significance of the present results for authoritarian personality theory depends rather critically on the status of one assumption: that the authoritarian mother's relative preference for nonlove-oriented punishment is a cause as well as an effect of authoritarianism. Assuming that the children of authoritarians themselves tend to be authoritarian, one can speculate about the role of avoidance tendencies in the etiology of authoritarianism. It may be that the child of authoritarian parents is highly ambivalent toward them (and other authorities) because of a heightened avoidance gradient, resulting in more closely balanced approach and avoidance tendencies. His later behavior toward superordinates typically reflects "coping techniques"—such as emphasis on the maintenance of social distance, interaction between formal roles rather than persons as individuals, etc.—designed to manage this ambivalence. But, as many writers on authoritarian personality have contended, there remains an underlying resentment and hostility toward authority that is revealed in such displaced forms as scapegoating and prejudice.

The speculative tenor of the above remarks

could be avoided if the child's actual behavior toward and perception of his parents, and reaction to different forms of socialization pressure, were determined by direct and indirect means in subsequent research.

#### SUMMARY

The primary purpose of this study was to investigate the way in which the mother's authoritarianism affects her selection of behavior control techniques in a variety of socialization situations. Specific predictions were formulated in an attempt to combine the hypotheses and insights of authoritarian personality theory (1) with the socialization model suggested by Whiting and Child (7). One hundred twenty-six mothers of preschool children responded in a structured interview to 38 behavior situations which commonly arise and call for maternal decisions about mode of control. These items were classified as relating to feeding, cleanliness and toilet training, sex, aggression, dependence, and independence; the preferred response of the mother could in turn be classified as love-oriented, nonlove-oriented or "ambiguous."

The main results of the study may be summarized as follows:

1. As predicted, authoritarians show a consistent tendency to select more nonlove-oriented and fewer love-oriented responses than nonauthoritarians. When the middle-range authoritarians are included, inspection indicates the relationship is linear.
2. There are significant variations in the proportion of love- and nonlove-oriented responses selected as a function of the behavior area involved. Thus, aggression and independence involve behaviors that provoke more nonlove-oriented responses, whereas sex and dependency behaviors call forth fewer nonlove-oriented responses.
3. There is no significant interaction, however, between authoritarianism and behavior area. That is, in terms of the a priori classification of items into six different areas, authoritarianism does not discriminate significantly better in one area than in another. It is quite possible, though, that different item arrangements (in terms more precisely relevant to authoritarianism) would show that the authoritarian's relative preference for nonlove responses is to some extent conditioned by the situation.



## REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper & Bros. 1950.
2. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281-302.
3. HART, I. Maternal child-rearing practices and authoritarian ideology. Unpublished doctor's dissertation, Duke Univer., 1956.
4. LEVINSON, D. J., & HUFFMAN, PHYLLIS E. Traditional family ideology and its relation to personality. *J. Pers.*, 1955, **23**, 251-273.
5. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.
6. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135-234.
7. WHITING, W. M., & CHILD, I. L. *Child training and personality: A cross-cultural study*. New Haven: Yale Univer. Press, 1953.

Received November 12, 1956.

# CATHARSIS AND THE RELIEF OF HOSTILITY

PHILIP WORCHEL

*University of Texas*

CATHARSIS, in one form or another, occupies a central position in all psychotherapeutic procedures. Corsini and Rosenberg (4), in reviewing the dynamics of group therapy, found that ventilation, or the release of feelings and expression of ideas usually repressed in other nontherapeutic situations, was one of the most frequently mentioned mechanisms in about 300 articles examined. Despite its importance, considerable confusion exists concerning the meaning and rationale for the efficacy of catharsis. Rotter (12) pointed out the many different ways in which catharsis is currently viewed and evaluated the different techniques of attaining catharsis from a learning point of view. Dollard and Miller (5) emphasized the importance of the events following the expression of feelings or past memories rather than the sheer act of catharsis. The therapist as a nonpunitive agent helps to extinguish certain avoidance reactions or to develop new learning by pointing out associations not originally perceived by the patient. Alexander and French (1) attributed the therapeutic outcome following catharsis to insight or new learning rather than to any purging effect of the unconscious. New learnings may also occur merely as a result of new relationships perceived by the patient as he verbalizes previous experiences.

Relatively few investigations have been undertaken to validate the clinical impression of the psychotherapeutic efficiency of catharsis or test any hypotheses as to its rationale. Haggard (7), using electric shock to induce disturbance, tested the relative efficacy of "catharsis-information" and experimental extinction or "rest" on reducing the general level of disturbance as measured by autonomic variables. Though he found that "catharsis-information" was significantly more effective, the design did not permit determination of whether "catharsis" or the straightforward "information" that shock would not be given was the essential variable. McKeachie *et al.* (8) found that students who were encouraged to write comments about their questions on

an examination made higher scores than students who had conventional answer sheets. The subjects, however, who were instructed to write explanations made slightly, but not significantly, better scores than those who were instructed to write feelings. The results raised doubts concerning the efficacy of "blowing-off steam." Wiener (13) investigated the relative effectiveness of two counseling techniques: reassurance-interpretation and catharsis-reflection. Though the combined counseled groups showed consistent but not always significant improvement in performance in comparison with the noncounseled groups, there were no indications that either of the two experimental counseling techniques was more effective than the other.

These ambiguous and contradictory results concerning the efficacy and the nature of effects of catharsis may be due to the untested assumption that catharsis in whatever form is equally applicable to all kinds of disturbances regardless of the underlying dynamics. It is suggested, however, that catharsis is effective to the extent that it sets up conditions for reducing threat to need-satisfaction patterns. Conflict may arise from a need to express hostility or a threat to self-esteem or between the moral values of the super-ego and the libidinal impulses of the id. Mere expression of anxiety feelings concerning an anticipated failure, for example, may not, in and of itself, eliminate the underlying threat. An analysis of the source of conflict in the expression of hostility suggests certain hypotheses concerning the role of various responses in the reduction of the conflict. The goal of the hostility drive is attack and removal of the stimulus arousing hostility. The threat to such direct behavior, however, is the anticipated punishment derived from earlier experiences in expressing hostility in this manner. It is hypothesized, therefore, that hostility will be reduced if the therapeutic situation provides socially acceptable conditions for the expression of hostility or attack at the instigator.

There has been some question (12), however,



concerning the generalization of the relief of such feelings when expression of hostility is made to a third party (e.g., the therapist) or displaced upon some object or person other than the original instigator. On the one hand, expression directly to the instigator, even under nonpunitive conditions, may still arouse anxiety concerning the anticipated consequences of such attack. On the other hand, a number of experiments (9, 10) have indicated the generalization effect occurring under conditions of displacement. A second objective, therefore, of the present series of studies is to compare the effect of expressing hostility directly to the instigator, and expressing hostility indirectly to or about others.

## METHOD

### *Subjects*

From nine sophomore and junior classes in elementary and educational psychology, 52 subjects (Ss) were selected from each of the three largest groups, 52 Ss from two smaller groups, and 52 Ss from the two smallest groups. The five groups were randomly assigned to the experimental treatments and were identified as follows: Group I, Catharsis-Experimenter (E); Group II, Catharsis-Assistant (A); Group III (E); Group II, Catharsis-Assistant (A); Group III (E); Catharsis-Fantasy (F); Group IV, Control (Talk); Group V, Control (Silence). Sex was about evenly divided in all groups: in Group I, there were 23 women and 29 men; in Group II, 25 women and 27 men; in Group III, 23 women and 29 men; in Group IV, 28 women and 24 men; and in Group V, 22 women and 30 men.

### *Experimental Tasks*

Basically, the design involved the arousal of hostility in groups of Ss, exposing them to different conditions of ventilating feelings, and then testing the efficiency of the experimental conditions in the reduction of the disruptive effects of stress on performance. Hostility was aroused by an insult technique similar to the one used by Feshbach (6) with one modification; insult was directed at the performance of Ss in an ego-involved task, an intelligence test. In the present experiment, catharsis refers to the release of hostility feelings, verbally expressed by volunteers in a group, under nonpunitive conditions. The tasks used to assess the degree of the reduction of stress were a digit-symbol test and a test of incidental recall of the titles of a bogus intelligence test. The basic assumption underlying the use of performance measures as indices of the disruptive effects of stress follows from the interference interpretation of the effects of frustration (2) by Child and Waterhouse (3): Responses such as anxiety and anger evoked by frustration in one activity may interfere with the quality of performance in a second activity.

### *Procedure*

To increase involvement, E introduced himself as professor of psychology and director of clinical psychology. Each group of Ss was told that as a part of the general routine of the Department of Psychology, students were given a series of tests measuring learning ability and intelligence. The scores were entered in each student's file for purposes of guidance, recommendations for scholarships and jobs, and consideration for admission to graduate study if the students applied for such admission.

As a pre-experimental test of learning ability, each group was given a one-minute digit-symbol test, the code consisting of letters and geometrical figures. Then a bogus intelligence test (14) entitled "The University of Texas Intelligence Test", and consisting of 14 subtests, each prominently labeled by a single word describing the contents, was administered to each group. The instructions, read verbally by E, indicated that this test consisted of a number of subtests, each highly related to academic success, and that the time limits of 45 to 90 seconds allowed sufficient time for most students to complete each subtest. At the beginning of each subtest, E said, "This test is called a test of \_\_\_\_\_. It is a measure of your ability to \_\_\_\_\_. Are you ready? Begin!" The time limits for the first two subtests permitted most of the students to complete them, but for the remaining 12 subtests, the material became much more difficult and the time for completion was sufficient for only a very few Ss. In order to arouse hostility feelings, E went around the room, critically remarking at the poor and slow performance of the Ss, urging more rapid solutions, and stating how much faster and better other classes were in this test.

The conditions during the period of eight minutes following the intelligence test varied for each group depending upon the experimental situations involved. In Group I (Catharsis-E), Ss were informed that there would be a slight rest period before the next test and that during this time the examiner would like to know how they felt about the test and the administration of the test. For Group II (Catharsis-A), E said that he would have to go to his office to bring more test material but that a teaching assistant would come in to take over until his return. The assistant, who had been instructed beforehand to profess ignorance concerning the events in the classroom, asked the group to inform him as to what had happened and to give him their reactions. The role of E or assistant for each group was merely to accept the statements without evaluation or comment except for a "hmm" or repetition of the statement. At the end of eight minutes, E walked in with more papers and continued the experiment. During the rest period for Group III (Catharsis-F), E mentioned that the university had surveyed student opinion a year ago concerning "irritations at the campus." As a follow-up on the previous survey, the class was asked to discuss their evaluation of the present situation.

In order to control the effect of "just talking" as against actually expressing hostility, a "neutral" topic was assigned for discussion in Group IV (Control-Talk). There had been recent publicity concerning the

shortage of science majors. The students were asked to suggest methods of encouraging more students in high schools to major in science. Finally, *E* lectured to Group V (Control-Silence) for eight minutes on the preparation of majors in psychology, allowing neither "just talking" nor the expression of hostility.

Following the "rest" period, each group was given another form of the digit-symbol test (numbers and different symbols). At the end of one minute, *Ss* were asked to write down the titles or brief descriptions of all the subtests in the intelligence test (incidental recall). All groups were tested within a two-day period to minimize communication to other *Ss*.

When all five groups had been tested, *E* described the nature of the experiment and full procedure with particular emphasis on the function of the bogus intelligence test to all *Ss* in order to relieve any doubt concerning their intellectual status.

## RESULTS

### *Expressions of Hostility*

A graduate assistant, unaware of the objectives of the present experiment, sat with the *Ss* in each group and made notes of their remarks during the experimental session. Though it is difficult to assign a global rating of the intensity of expression for the group as a whole, the nature of the remarks indicated rather clearly that reports of expressed hostility were more frequent and intense in Groups II and III than in Group I. Even though *Ss* in Group I were encouraged to express their feelings, they gave mild criticisms only of the test. They suggested that the time limits were too short, *E* did not give sufficient time for them to read instructions, the items were too difficult, and the test was not a valid measure of their intelligence. One might infer that hostility, if present, was directed toward the test. In Group II, *Ss* talked more directly about *E* to the assistant. They said he "talked too much," "he was too cocky," "who did he think he was," "what the hell was his name,"

"if he kept his mouth shut they could have done better." Group III was very critical of staff members and teaching assistants: "Professors rewrote their books every year so that students have to buy new textbooks," "professors were often unprepared for their lectures," "some had no business to teach," "why don't professors stay in their offices to see students," "teaching assistants gave unfair exams and didn't know any more than the students." *Ss* seemed able to report more hostile feelings about the *E* to somebody else than to the *E* directly. When they had to express their feelings directly to *E* in Group III, however, they expressed more hostility about other authority figures (professors and assistants) than about *E* himself. In the control situation of Group IV, *Ss* suggested methods of encouraging more science majors such as increased publicity, more use of audiovisual techniques in science classes, and emphasis upon applications of science courses. In addition, however, there were a few critical remarks on "poor science teachers" and "inadequate pay for good teachers."

### *Learning*

It was predicted that hostility feelings would be reduced under conditions encouraging expression of attack or hostility. Specifically, therefore, scores on the digit-symbol test should be higher for the three experimental than for the two control conditions.

For the digit-symbol test, Table 1 shows the pre-experimental and postexperimental means for all five groups. Since the pre-experimental means differed in the five groups, an analysis of covariance was applied. The adjusted mean scores are shown in the last column.

Since an obtained *F* ratio of 11.43 for the treatment variable was significant well beyond the .01 level, tests of significance among the five groups of *Ss* were performed and are reported in Table 2. Since adjusted mean digit-symbol scores for Groups I, II, and III are significantly higher (.01 level) than those for Groups IV and V, the hypothesis seems clearly supported. There is no significant difference, however, among the adjusted means of Groups I, II, and III. Nevertheless, Group IV, which merely discussed a neutral topic,

TABLE 1

ADJUSTED POST-TREATMENT LEARNING SCORES ON THE DIGIT-SYMBOL TEST FOR THE FIVE GROUPS

Group	Observed Mean		Adjusted Mean ( $\bar{Y}$ )
	$\bar{Y}_1$	$\bar{X}_1$	
I Catharsis-E	44.96	28.44	44.52
II Catharsis-A	46.15	28.90	45.34
III Catharsis-F	44.48	27.69	44.65
IV Control (talk)	42.61	28.17	42.37
V Control (silence)	38.83	26.11	40.24



TABLE 2

TESTS OF SIGNIFICANCE BETWEEN THE ADJUSTED MEANS OF THE DIGIT-SYMBOL SCORES FOR THE FIVE GROUPS

Groups	Adjusted Mean Diff.	<i>t</i> Ratio	<i>P</i>
I and II	.82	.96	—
I and III	.13	.15	—
I and IV	2.15	2.43	.01-.02
I and V	4.28	4.84	<.01
II and III	.69	.78	—
II and IV	2.97	3.36	<.01
II and V	5.10	5.76	<.01
III and IV	2.28	2.58	<.01
III and V	4.41	4.98	<.01
IV and V	2.13	2.41	.01-.02

TABLE 3

TESTS OF SIGNIFICANCE BETWEEN THE MEANS OF THE FIVE GROUPS ON THE POST-TREATMENT TEST OF INCIDENTAL RECALL

Groups	$\bar{X}_D$	<i>t</i> Ratio	<i>P</i>
I and II	1.21	3.36	<.01
I and III	.77	2.14	.02-.05
I and IV	.68	1.95	.05
I and V	.85	2.36	.01-.02
II and III	.44	1.22	—
II and IV	1.89	5.22	<.01
II and V	2.06	5.72	<.01
III and IV	1.45	4.00	<.01
III and V	1.62	4.50	<.01
IV and V	.17	.50	—

did significantly better (.01-.02 level) than Group V, where the Ss maintained silence.

#### Incidental Recall

The results of the test for incidental recall not only confirm the findings on the digit-symbol test of significantly superior performance by the three experimental over the two control groups, but also point to a significantly superior performance by Groups II and III over Group I. An obtained *F* ratio of 12.48 was significant beyond the .01 level for the treatment variable. As shown in Table 3, the mean incidental recall scores (number of titles recalled) of 6.50 and 6.06 for Groups II and III, respectively, are significantly higher than the mean of 5.29 for Group I. All three mean scores for the experimental groups are significantly higher than those of 4.61 and 4.44 for the two control groups (Groups IV and V, respectively). No difference between the mean scores of Groups II and III or between the two control groups was demonstrated. Thus, incidental recall follow-

ing expression of hostility to the instigator about other authority figures or following expression of hostility about the instigator to another authority figure seems to have a significantly more facilitative effect on performance than the direct expression of hostility to the instigator.

#### DISCUSSION

If the learning and incidental recall indices can be accepted as indirect indicators of the effects of the relief of aggressive tension, then the results clearly indicate (a) the therapeutic effects of catharsis on the relief of aggressive tension, and (b) the differential effects of varying cathartic conditions. It may be that hostility was relieved by ventilation or that expression of hostility about *E* or other similar figures in the cathartic situation permitted a reduction of hostility feelings towards *E* in the postexperimental test situation (generalization). Certain limitations in the present experiment, however, should be recognized. Since groups rather than individuals were employed in the study, relatively few Ss reported their feelings during the brief cathartic period. It may be that the silent members relieved hostility feelings vicariously as others expressed aggression, or the effect may have been largely restricted to the more verbal members. The experiment warrants replication so that the differential effects on the verbal and silent members may be studied.

Another factor to be considered in interpreting the results is the brief and "single-shot" nature of the cathartic period. It is possible that hostility would be more extensively relieved in Group I (direct expression to *E*) if a number of cathartic sessions were used. A number of trials might be necessary for Ss in Group I before the free expression of hostility feelings could occur. Though the results indicate that hostility feelings were reduced as much if not more in Groups II and III than in I, nothing can be said as to the duration or generalization of these feelings (postexperimental sessions). Experiments on displacement (9, 10, 11) suggest that the approach gradient would rise for the original goal after approach to similar stimuli. Brief encounters with Ss in the corridors after the testing of the initial groups and

prior to the explanation of the experiment suggest that hostility was still strong towards *E*. In the present experiment, the conditions in Groups II and III might be conceived as involving displacement of hostility feelings. In one case, the object of attack, *E*, was kept constant, but the person to whom the feelings were expressed varied from *E* (Group I) to the assistant (Group II). In the other case, the feelings were expressed to the same person, *E*, but the object of attack varied from *E* (Group I) to other staff members (Group III). Traditionally, the latter case is more appropriately an example of displacement since the object of attack is similar but not identical to the original instigator. The former case, more like the therapeutic situation, may nevertheless be considered an example of displacement since the punitive agent to whom hostility is expressed differs from the original.

One final consideration involves the role of *E* in the two control groups. *E* had to be present in order to equate the condition of his presence in Groups I and II, but a more complete design to test the effect of the assistant in Group II should include a replication of the two control groups with the assistant rather than *E* in charge of the eight-minute control period. It is possible that the lecture by *E* in Group V might have heightened temporarily any feelings of hostility, or that discussion concerning increasing the number of science majors in Group IV may have reduced hostility feelings through some critical comments about poor teaching. Either situation, the lecture by *E* in Group V or the discussion of science majors in Group IV, might explain the significantly better results for Group IV on the digit-symbol test.

#### SUMMARY

The purpose of the present study was to study the effect of different cathartic techniques on the relief of hostility feelings. From nine sophomore and junior classes in elementary and educational psychology, 52 *Ss* were selected from each of the three largest groups, 52 *Ss* from two smaller groups, and 52 *Ss* from the two smallest groups. All groups were given a bogus intelligence test consisting of 14 subtests to produce ego-involvement. Hostility was aroused by the use of "insult" during the administration of

the tests. During the "cathartic" sessions of eight minutes, Group I was asked to express criticism of the administration of the test. For Group II, *E* mentioned that he was going to his office for other test material and that an assistant would take over for the next few minutes. The assistant professed ignorance of the proceedings and asked *Ss* to describe the previous events and state their reactions to them. Group III was requested to give a follow-up evaluation of a study made a year ago on campus irritations. Group IV, a control group, was asked to discuss a "neutral" topic, and V, another control group, was not permitted to say anything while *E* lectured on the preparation of majors for psychology. A pre- and post-treatment digit-symbol test and an incidental recall test of the 14 titles of the subtests of the intelligence test were used as measures of the effect of the relief of hostility feelings. The results were as follows:

1. *Ss* in Groups II and III were able to express more hostile feelings than those in Group I, who were asked to express feelings about *E* directly to him.
2. Adjusted digit-symbol scores were significantly higher for the three "cathartic" groups (I, II, III) than for the two control groups (IV, V).
3. There were no significant differences in the adjusted mean digit-symbol scores among the three experimental groups.
4. On incidental recall, *Ss* in Groups II and III were significantly higher than all other groups.
5. *Ss* in Group I gave significantly higher incidental recall scores than those in the control groups (IV and V).

#### REFERENCES

1. ALEXANDER, F., & FRENCH, T. M. *Psychoanalytic therapy: Principles and application*. New York: Ronald Press, 1946.
2. BARKER, R. G., DEMBO, TAMARA, & LEWIN, K. Frustration and regression: An experiment with young children. *Univ. Iowa Stud. Child Welf.*, 1941, 18, 1-314.
3. CHILD, I. L., & WATERHOUSE, J. K. Frustration and the quality of performance: I. A critique of the Barker, Dembo, and Lewin experiment. *Psychol. Rev.*, 1952, 59, 351-362.
4. CORSINI, R. J., & ROSENBERG, B. Mechanisms of group psychotherapy: Processes and dynamics. *J. abnorm. soc. Psychol.*, 1955, 51, 406-411.
5. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.



6. FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, 50, 3-11.
7. HAGGARD, E. A. Some conditions determining adjustment during and readjustment following experimentally induced stress. In S. S. Tomkins (Ed.), *Contemporary psychopathology*. Cambridge, Mass.: Harvard Univer. Press, 1947. Pp. 529-544.
8. McKEACHIE, W. J., POLLIE, D., & SPEISMAN, J. Relieving anxiety in classroom examinations. *J. abnorm. soc. Psychol.*, 1955, 50, 93-98.
9. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, 43, 155-178.
10. MURRAY, E. J., & BERKUN, M. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.*, 1955, 51, 47-56.
11. MURRAY, E. J. A case study in a behavioral analysis of psychotherapy. *J. abnorm. soc. Psychol.*, 1954, 49, 305-310.
12. ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
13. WIENER, M. The effects of two experimental counseling techniques on performance impaired by induced stress. *J. abnorm. soc. Psychol.*, 1955, 51, 565-572.
14. WORCHEL, P. Anxiety and repression. *J. abnorm. soc. Psychol.*, 1955, 50, 201-205.

Received November 30, 1956.

# ASSIMILATION AND CONTRAST EFFECTS IN REACTIONS TO COMMUNICATION AND ATTITUDE CHANGE<sup>1</sup>

CARL I. HOVLAND

*Yale University*

O. J. HARVEY

*Vanderbilt University*

AND

MUZAFER SHERIF

*University of Oklahoma*

THIS paper presents an experiment on reactions to communication and on attitude changes by individuals whose initial stands on a controversial social issue diverged in varying degrees from positions advocated in communication. Study of the relationship between subject's (S's) attitude and the position advocated in communication may help resolve some apparently contradictory effects of communication aimed at changing attitudes.

Attempts to change attitudes in the direction advocated by communication on a social issue at times produce shifts in the direction opposite to that intended—the "boomerang effect." While numerous investigators have reported shifts of average test scores in the direction of communication (4), a fairly common finding, even in these studies, is that some individuals shift their stand *away* from that presented in communication. Several studies reporting both positive and negative shifts in attitudes toward out-groups following communication are summarized by Williams (20).

Thus, at times, persuasive communication produces a bi-modal distribution of attitude scores (11, pp. 874-875). For example, Remmers (12) obtained positive shifts on average scores following communications on conservation, social insurance and labor unions, but the latter communication "sharply divided

the group into two opposing tents" (12, p. 201). In Knower's study prior to repeal of prohibition, "wet" communication to generally "dry" Ss and "dry" communication to generally "wet" Ss resulted in shifts in both positive and negative directions in each group (8). A rather striking instance of such opposing effects was reported by Wilke (18), whose antiwar communication was presented at a time when the student population from which Ss were drawn was divided in controversy over this very issue.

A few studies, such as those by Manske (10) and Russell and Robertson (14), have found group shifts in the direction opposite to the stand presented in communication. A related finding reported by Williams and Remmers (19) following communication on a rural issue was reduced variability and a comparatively less favorable stand by a group of rural youth in contrast to the increased variability in an urban group. Some authors who have obtained results in the direction away from communication insert the suggestion that too great divergence between S's stand on the issue and the stand presented in communication may have been responsible.

On the other hand, data from several recent studies suggest that the extent of influence increases as a function of the distance between position of communication and position of the recipient. Goldberg (2) reports that the greater the discrepancy between an announced group norm and the S's own judgment the greater the change produced. In a still unpublished paper, French and Gyr (summarized in 1) found a positive correlation between the degree of deviation between inducer and inductee and the amount of change. In a perceptual task, Lubin and Fisher (9) obtained an increase in conformity to a partner's judgment as the distance increased up to a point where with great distance the proportion of movement declined. Hovland and Pritzker (6) have shown that the larger the change in opinion advocated the greater the amount of

<sup>1</sup> This investigation was conducted as part of the Yale Communication Research Program. This series of studies, devoted to an analysis of factors related to attitude and opinion formation, is financed by a grant to Carl I. Hovland from the Rockefeller Foundation, whose support is gratefully acknowledged. Thanks are extended to Dr. Charles Shedd, now of Berea College, Mr. William Smith, and Mr. Richard Disney who participated in developing the procedures and in writing the communications used. We are also grateful to Dr. Laurence H. Snyder, Dean of the Graduate College, University of Oklahoma, for necessary administrative arrangements, and to Drs. Irving L. Janis, Robert P. Abelson, and Jack W. Brehm for suggestions made in the course of reading the manuscript.



change produced. The issues that were employed did not pertain to strongly rooted attitudes, but represented opinions on rather factual topics. The communicators used were authorities respected by the recipients.

A comprehensive analysis may be necessary to handle the above findings as well as results of studies employing communication on controversial social issues which suggest that too great distance between *S*'s attitude and the position advocated in communication produces "boomerang effects." One possible approach is through analysis of processes underlying judgments of motivationally neutral material, exemplified by psychophysical stimuli, as well as ego-involving verbal material, exemplified by controversial social issues. Investigations of judgmental processes and their theoretical relevance for the study of reaction to communication and attitude change will be presented in a forthcoming volume.

One set of findings suggests that *S*'s position shifts toward the stand advocated in communication when the topic is not highly ego-involving and *S*'s position does not diverge in the extreme from the stand advocated. This result seems akin to extension of an established reference scale in judgment of weights or inclinations following introduction of anchoring stimuli near the end stimulus of the series, as reported by Rogers (13), Heintz (3), and others. Sherif, Taub, and Hovland (16) demonstrated this extension toward the anchor ("assimilation effect") as well as the effect of more remote anchors beyond either end of the stimulus series in constricting the reference scale of judgment and displacing judgment away from the anchor ("contrast effect"). Contrast phenomena may also appear in reaction to communication on an issue which is not highly involving for *S* if the position advocated is removed sufficiently from *S*'s position, as the Lubin and Fisher results seem to indicate.

The above analysis based on findings from judgment studies suggested the further possibility that when *S*s have established attitudes and are personally involved in a controversial social issue, their "own stand" functions as the major anchorage affecting reaction to and evaluation of communication. In this case, communication near *S*'s stand would be assimilated to it, while communication at

variance with *S*'s own stand would be displaced still further away ("contrast effect"). Whether assimilation or contrast effects appear would be a function of the relative distance between *S*'s own stand and the position of communication.

Accordingly, in the present experiment communications representing two opposite extremes and one moderate position on an ego-involving issue were presented to *S*s whose initial stands on the issue ranged from one extreme to the other. The following hypotheses were formulated in terms of the effect of the relative distance between *S*'s own stand and the position advocated upon evaluation and placement of communication as well as acceptance-rejection of that position:

1. Reactions to a communication will decrease in favorableness as the distance between *S*'s own stand and the position advocated in the communication increases.

2. In evaluations by *S* of what position is advocated by a communication, the greater the distance between *S*'s own stand and the position advocated in the communication, the greater the displacement *away* from *S*'s position ("contrast effect"). When only a small discrepancy in position exists there will be a tendency for displacement *toward* *S*'s stand ("assimilation effect").

3. With small distances between the position of the communication and that of the *S*, changes in *S*'s opinion in the direction advocated by the communication will occur. With large distances between the stands taken by communication and by *S*, opinion changes in the direction advocated will be infrequent.

## METHOD

### *Issue and Communications*

The problem required that a controversial issue be selected, that *S*s' stands on the controversy be ascertained, and that communications advocating various positions be presented to *S*s with differing stands on the issue.

The issue chosen was the controversy over prohibition and repeal in a "dry" state. Shortly before the study began, a referendum was held to determine the fate of existing prohibition laws. The vote favored prohibition by a narrow margin.

In order to differentiate existing stands on the issue, representative statements made during the referendum campaign were collected from leading newspapers in two large cities. In addition, statements were obtained from 500 people in several localities on a random basis. These statements from public and private sources were sorted by 20 judges to secure clearly differentiated

stands actually taken in the state. As a result, eight statements were chosen representing prevailing stands ranging from strong advocacy of prohibition to strong advocacy of repeal. One additional "wet" statement was added as a logical counterpart of the most extreme "dry" stand, giving a total of nine statements. These statements are as follows, (I) being the additional item:

(A) Since alcohol is the curse of mankind, the sale and use of alcohol, including light beer, should be completely abolished.

(B) Since alcohol is the main cause of corruption in public life, lawlessness, and immoral acts, its sale and use should be prohibited.

(C) Since it is hard to stop at a reasonable moderation point in the use of alcohol, it is safer to discourage its use.

(D) Alcohol should not be sold or used except as a remedy for snake bites, cramps, colds, fainting, and other aches and pains.

(E) The arguments in favor and against the sale and use of alcohol are nearly equal.

(F) The sale of alcohol should be so regulated that it is available in limited quantities for special occasions.

(G) The sale and use of alcohol should be permitted with proper state controls, so that the revenue from taxation may be used for the betterment of schools, highways, and other state institutions.

(H) Since prohibition is a major cause of corruption in public life, lawlessness, immoral acts, and juvenile delinquency, the sale and use of alcohol should be legalized.

(I) It has become evident that man cannot get along without alcohol; therefore, there should be no restriction whatsoever on its sale and use.

Three communications of equal length were prepared, each requiring approximately 15 minutes for delivery. Arguments were those actually made by prohibition and repeal advocates during the referendum campaign. Arguments in the three communications were arranged in the same order, but from the viewpoint of the different parties to the controversy. One communication presented an extreme "dry" stand; one an extreme "wet" stand; and one a moderately wet stand, as typified in statement (F) above. All communications were presented by tape recording. The same voice was used in recording wet and dry communications.

### Procedure

In the first session, data on Ss' attitudes were obtained. At the time, Ss were not told that an additional session would be held. A schedule on "public issues" was presented with assurance of anonymity to Ss. Following a "dummy" issue (college football) Ss responded to the nine representative statements on prohibition. The following instructions were printed on the schedule and read aloud by the experimenter (E):

Below are some statements recently made concerning the wet-dry issue in this region.

Please read *all* of the statements carefully first before making any marks on this page.

Now that you have carefully read all the state-

ments, underline the *one* statement that comes closest to your own point of view on the topic.

There may be other statement or statements which you find not objectionable from your point of view. Put a circle around the letter in front of such a statement or statements which are *not objectionable* to you.

Now cross out that one statement which is *most objectionable* from your point of view.

There may be other statement or statements which you find *objectionable* from your point of view. Cross out the letter in front of such a statement or statements which are *objectionable* to you.

These procedures yielded data on S's stand and also on the range of his tolerance for other stands in the series (*latitude of acceptance*), the range of stands he rejected (*latitude of rejection*) and those stands which he did not consider either acceptable or unacceptable. Thus the procedures gave S an opportunity not to take a stand in relation to items which he did not include in his latitude of acceptance or latitude of rejection, rather than requiring artificial "indifference" or "neutral" checkings.

From 1-3 weeks after the first session, a communication was presented by tape recording. Arrangements for its presentation were made through a member of the group being exposed and it was introduced as a talk actually made by a proponent of the stand advocated. The wet (repeal) communication was presented to extreme dry Ss and unselected Ss. The dry (prohibition) communication was presented to extreme wet and unselected Ss. The moderate communication was presented to wet, dry, and unselected Ss. The Ss participated in both "before" and "after" sessions in small groups of 10-30 under close supervision.

Following the communication, the same questionnaire for securing S's stand, latitudes of acceptance and rejection was filled in a second time. Reactions to the communication presented were obtained through ratings on like-dislike, reasonable-unreasonable, biased-unbiased, propaganda-fact dimensions. The Ss in the moderate communication groups also checked on a graphic rating scale ranging from extreme dry (A) to extreme wet (I) positions what they thought to be the stand taken in the communication itself.

### Subjects

Since the objective was to secure Ss who were definitely ego-involved, a special point was made of obtaining Ss with established and publicly committed dry or wet stands as a validity check of the paper-and-pencil checkings. It was not difficult to select Ss in the dry groups on the basis of known information concerning their stand. Two small samples from Women's Christian Temperance Union groups were obtained and a group of Salvation Army workers. The four remaining dry groups were students in preparation for the ministry or in strict denominational colleges. A total of 183 dry Ss participated in both sessions. It was much more difficult to obtain Ss whose position was known to be wet. However, on the basis of cases personally known to the Es or their assistants, 25 wet Ss were secured. For comparison, 290 additional Ss were obtained representing more moderate positions



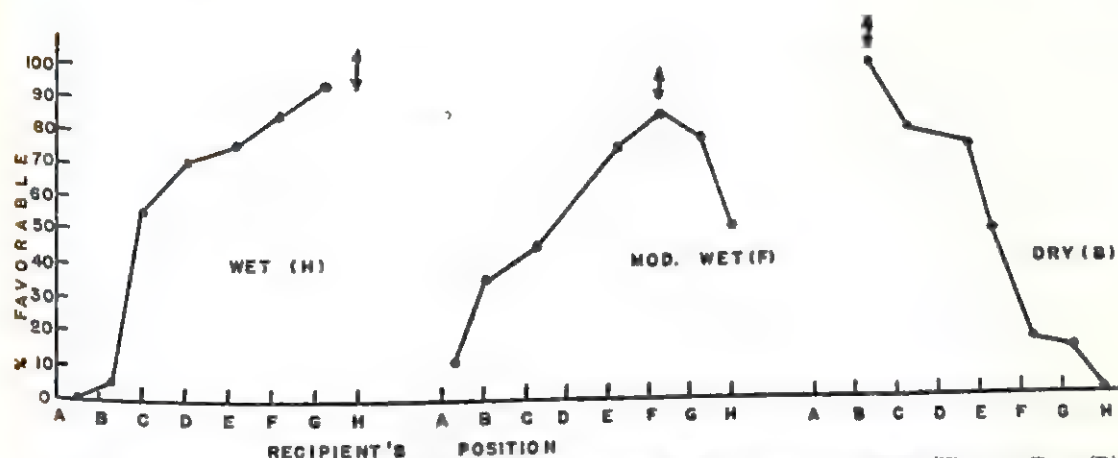


FIG. 1. PERCENTAGE OF FAVORABLE EVALUATIONS OF WET (H), MODERATELY WET (F), AND DRY (B) COMMUNICATIONS FOR Ss HOLDING VARIOUS POSITIONS ON PROHIBITION (BASED ON MEAN ACCEPTABLE STATEMENT)

(Positions of communications indicated by arrow)

on the issue. These were college students secured from classes in journalism, speech, education, chemistry, etc. All Ss were residents of the dry state where prohibition was a lively topic of controversy at the time of the study. For the reason indicated, it was not possible to match the age levels of Ss in the three categories.

## RESULTS

### *Evaluation of the Communication*

A five-item scale was used to measure the audience's evaluation of the fairness and impartiality of the communication. In Figure 1, the percentages of Ss in each group whose reactions were favorable are graphed for the three communications. On the abscissa the stands of the recipient (mean of statements checked acceptable) are presented and on the ordinate the percentage of favorable evaluations. It will be observed that there is an extremely close relationship between the individual's own stand on the issue and his evaluation of communication. The two communications advocating extreme positions have their peak of favorable responses among those holding corresponding extreme positions. The maximum favorable reaction for the moderate communication is found among those holding a moderate position. The data provide quantitative information to support the expectation from earlier studies (5, 17) that individuals who are in favor of the opinion advocated will consider the communication fair and unbiased, but that those with an opposed stand will regard an identical communication as propagandistic and unfair.

### *Placement of the Position of the Communication*

The principal results of the experiment are those pertaining to the recipient's perception of the stand advocated in the communication. We asked S to indicate on a graphic scale ranging from extreme dry position to extreme wet position, what he thought to be the position of the moderately wet communication (at F). From our previous study on assimilation-contrast effects with psychophysical data we predicted that positions differing only slightly from one's own would be "assimilated," while larger differences between one's own position and that of the communication would be exaggerated, showing a "contrast" effect. In Figure 2, results concerning this prediction are presented. The S's own position is indicated on the abscissa and along the ordinate the average rating of the position of the communication is given for the speech in which a moderately wet position (about F) was advocated. The dots indicate the mean placement of the position of the communication for Ss who indicate each particular position as their own stand. The squares represent the mean placement of the position advocated when S's position is estimated from the mean of the acceptable positions checked. The dotted line indicates a hypothetical relationship in which Ss holding the same position as the communication report its position accurately (at F), those a small distance removed assimilate it to their own position, and those still further removed exaggerate the position revealing a contrast effect.

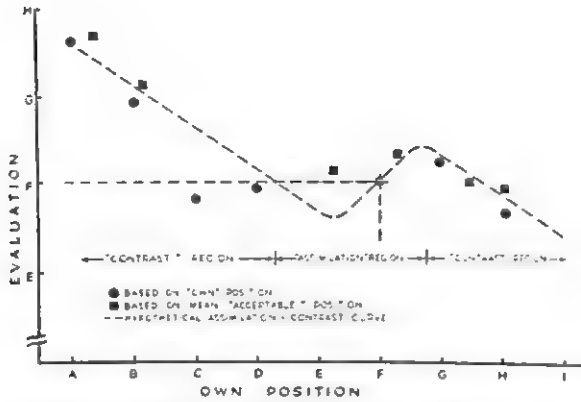


FIG. 2. AVERAGE PLACEMENTS OF POSITION OF MODERATELY WET COMMUNICATION (F) BY Ss HOLDING VARIOUS POSITIONS ON THE ISSUE, PLOTTED AGAINST HYPOTHETICAL ASSIMILATION-CONTRAST CURVE

The expectation is fulfilled that those at the wet end judge that the communication advocated a drier position than it did and those at the dry end judge that it was advocating a wetter position than it did. Those nearer the position of the communication reported it more objectively. The evidence concerning assimilation deviates somewhat from the theoretical curve. There were too few Ss with intermediate positions to enable us to determine these positions adequately.

#### Acceptance-Rejection of Stands

Each S was asked to indicate not only the stand which best represented his, but also other stands which were more or less acceptable (i.e., his latitude of acceptance), the stand most clearly opposed to his position, and others not acceptable (i.e., his latitude of rejection). From these data, we find that Ss with more extreme positions tend to reject more frequently positions not within their latitude of acceptance. The Ss with middle positions are more apt to rate positions removed from their own as indifferent. In Table 1, data are presented showing the mean number of items rated "also acceptable," "unacceptable" and "neither acceptable nor unacceptable" by Ss with strong positions (A, B, G, H, I) as compared with those by Ss with more moderate positions (C, D, E, F). It will be observed that a significantly greater number of items are judged unacceptable by Ss with extreme stands. Smaller differences between the groups exist in the number rated acceptable. These results tie in with our

TABLE 1  
ACCEPTABILITY OF STATEMENTS IN RELATION TO EXTREMITY OF SUBJECTS' POSITION ON ISSUE

S's Positions	N	Mean Number of Items Acceptable	Mean Number of Items Not Checked	Mean Number of Items Rejected
1. Extreme (A, B, G, H, I)	193	2.81	1.48	4.71
2. Intermediate (C, D, E, F)	37	3.05	2.24	3.70
<i>p</i>				< .03

earlier studies (7, 15) in which we found that Ss with extreme positions and considerable ego-involvement in an issue have raised thresholds of acceptance and lowered thresholds of rejection in placing items concerning that issue, resulting in smaller latitudes of acceptance and greater latitudes of rejection than for Ss with moderate positions. These findings can be used in constructing the expected latitudes of acceptance and rejection of the individual for an ego-involving issue once his own position is ascertained and possibly in predicting evaluation of communication which stands at a given distance from his own position. An attempt in this direction is presented in the discussion.

#### Changes in Opinion for Groups with Different Stands on the Issue

Finally, it would be expected on the basis of the considerations previously discussed that those holding a position at great variance with that being presented in the communication would be relatively little modified in their opinion. The index of opinion change was based on the mean position of the statements checked by Ss as being "acceptable" to them. A value of 1 indicates the extreme dry end in which only statement A was checked. Correspondingly, a value of 9 is at the wet end with only statement I checked. The positions of the three speeches at B, F, and H would thus be represented by values of 2, 6, and 8, respectively.

Table 2 gives the mean position for five criterion groups before and after the communication, together with the average amount of change. Only the unselected group exposed to the wet communication should be expected to show sizeable change. The four other groups all have initial positions quite removed from the position being advocated. While the



TABLE 2  
OPINION CHANGE  
Changes in Mean Acceptable Statement for Ss  
with Differing Initial Stands

Group	N	Before Comm.	After Comm.	Change in Direction of Comm.
Wet Communication (H)				
Drys	69	2.39	2.34	-.05*
Unselected	92	5.10	5.65	+.55*
Dry Communication (B)				
Wets	25	6.70	6.74	-.04
Unselected	87	5.90	5.78	+.14
Moderately Wet Communication (F)				
Drys	114	2.17	2.26	+.09

\* Difference between changes:  
 $p = <.03$  (one tail).

TABLE 3  
OPINION CHANGE  
Percentage of Ss Changing in Direction of  
Communication or in Opposed Direction

Group	N	Change in Direction of Comm.	No Change	Change in Di- rection Op- posed to Comm.	Net Change
Wet Communication (H)					
Drys	69	% 27.5	% 49.3	% 23.2	% +4.5*
Unselected	92	52.2	23.9	23.9	+28.3*
Dry Communication (B)					
Wets	25	24.0	56.0	20.0	+4.0
Unselected	87	40.2	33.4	26.4	+13.8
Moderately Wet Communication (F)					
Drys	114	31.6	49.1	19.3	+12.3

\* Difference between changes:  
 $p = <.04$  (one tail).

dry and unselected groups exposed to the wet communication do differ substantially in their initial positions, the positions of the wet and unselected groups are similar, since the unselected group (college students) initially holds a rather wet position.

As predicted, there is a significantly greater change for the unselected group given the wet communication than for the dry group ( $p = <.03$ ). As in all studies in which the variable with which we are concerned cannot be directly manipulated, the possibility must be considered that there are factors other than position on the issue which are correlated with it and provide the basis for the relationship obtained.

In Table 3, results are presented for the same groups in terms of the percentage of Ss who are influenced either in the direction advocated by the communication or in the direction opposed to it. The net change column represents the differences between those changed in the direction advocated and those changed in the opposite direction. The net change is greater among the unselected group than among the drys when a wet communication is presented ( $p = <.04$ ). Twenty-eight per cent of the former and only four per cent of the latter changed in the direction advocated. It will also be seen that the amount of change produced among the drys by a moderately wet communication (at F) is greater than that produced by the more extreme wet communication (at H). The net proportion changed by the former is 12 per cent, while it is only 4 per cent for the latter. This difference, however, is not statistically significant and, in any case, it is difficult to establish that the two different communications were inherently equivalent in persuasiveness.

It will be noted that by confining our comparisons to groups who were initially selected as having known differences in their stand on the issue (either wet or dry) we were able to avoid the regression effects often found in studies where amounts of change are compared for groups differentiated solely on the basis of initial scores on the attitude measure. Under the latter conditions, those with more extreme scores often regress toward the mean and this can either obscure the effect or produce a relationship artifactually.

Finally, it will be noted from Table 3 that the predominant response among Ss holding extreme positions is to remain uninfluenced by the communication. Even where Ss in these groups were influenced, changes were seldom found in the item they checked as best representing their own position. Instead, there was typically only an increase in the number of other statements intermediate between their own position and that of the communication which they would check as "also acceptable."

#### DISCUSSION

The results presented lend support to the three hypotheses pertaining to evaluation and placement of communication and attitude

TABLE 4

## HYPOTHETICAL LATITUDES OF ACCEPTANCE AND REJECTION OF Ss HOLDING EACH POSITION

(Columns show latitude of acceptance (strongly accept plus accept), latitude of rejection (strongly reject plus reject) and positions neither acceptable nor unacceptable to Ss holding given positions.)

Rating Positions	Own Position								
	A	B	C	D	E	F	G	H	I
A	✓✓	✓	0	×	×	×	×	×	×
B	✓	✓✓	✓	0	×	×	×	×	×
C	0	✓	✓✓	✓	0	×	×	×	×
D	×	0	✓	✓✓	✓	0	×	×	×
E	×	×	0	✓	✓✓	✓	0	×	×
F	×	×	×	0	✓	✓✓	✓	0	×
G	×	×	×	×	0	✓	✓✓	✓	0
H	×	×	×	×	×	0	✓	✓✓	✓
I	×	×	×	×	×	×	0	✓	✓✓

Code: ✓✓ = strongly accept; ✓ = accept; 0 = neither accept nor reject; × = reject; ×× = strongly reject.

Mean frequencies in above table:

	A,B,G,H,I	C,D,E,F
✓ and ✓✓	2.7	3.0
0	1.3	2.0
×	5.0	4.0

change as a function of the relative distance between S's stand and the position of communication on an ego-involving issue. Our findings suggest that the relative distance between Ss' attitudes and position of communication may be useful in explaining apparently contradictory effects of communication in producing attitude change in the intended direction, no change, and change in the opposite direction. This approach seems to have predictive value for S's reactions to other positions on an issue and susceptibility to change once his own stand is ascertained.

On the basis of results in Table 1 and evidence from previous studies (7, 15), we can formalize the pattern of acceptance and rejection for various positions by Ss holding each initial position. In Table 4, the expected response to each opinion item is given for Ss checking positions from A to I as their most acceptable stand. The S's own stand is indicated by ✓✓, "also acceptable" with ✓, completely unacceptable with ××, other stands unacceptable with ×, and not clearly acceptable or unacceptable with 0. Thus, the table presents a hypothetical pattern for the "latitude of acceptance" (consisting of S's own position and other acceptable positions) and "latitude of rejection" (consisting of all unacceptable positions) for Ss holding each

position. Where sufficient numbers of Ss are available for a given stand, the empirical distributions of responses show a close correspondence to these hypothetical patterns.

This model for latitudes of acceptance and rejection can be used to explicate results on reaction to communication. By comparison with Figure 2, showing placement of the position advocated in the moderate communication, it can be seen that "contrast effects" begin to appear when the position advocated falls within S's latitude of rejection, an area far removed from his own stand. If the position advocated in communication falls at the limits or slightly beyond the latitude of acceptance, it would be more likely to be assimilated to the latitude of acceptance, although this trend is less clear in the present study.

If we assume that evaluation of communication on an issue in which S is personally involved depends on two factors: (a) what the position of communications is perceived to be and (b) what segment of the acceptance-rejection range is associated with the perceived position, we can predict from Fig. 2 and Table 4 the degree of favorableness of evaluations of the communication (Fig. 1). When the communication falls within the latitude of acceptance (✓ or ✓✓), it is judged fair and unbiased. As the perceived position of communication moves outside the latitude of acceptance into the latitude of rejection (×), it is increasingly considered unfair and propagandistic. For example, an individual checking B as his most acceptable stand displaces the F communication to G ("contrast effect") which, according to Table 4, is clearly in his latitude of rejection; an S checking A as his own stand places the F communication between G-H, still further toward the extreme end of the latitude of rejection, and the percentage of favorable evaluations of the communication declines still further (cf. Fig. 1).

Two general problems are suggested by the present analysis. Since S's stands on an issue of personal concern produce variation in reaction to communication and attitude change, the effect of the degree of ego-involvement in issues needs exploration. It is expected that latitudes of acceptance and rejection will vary as a function of the degree of ego-involvement, and these variations may affect the



ranges in which assimilation and contrast effects in judging communication occur. A person strongly committed on an issue will be more discriminating in accepting stands (raised threshold of acceptance) while one who is less committed will be willing to consider a larger number of alternative positions (lowered threshold of acceptance). Corresponding effects on extent of opinion change with different distances between *S* and communication are to be expected. With low involvement issues one would anticipate increase in opinion change with increased separation over a considerable range, whereas with high involvement issues one would expect opinion change over only a narrow range of separation, with resistance to acceptance of the communication for the remaining distances.

Finally, an interesting problem remains as to the role of the communicator in influencing the range of assimilation. In the Hovland-Pritzker study (6) the communicator was an authority respected by *Ss* on the issue presented. There the greater the distance between communicator position and that of *S*, the greater the opinion change. In the present study, on the other hand, the communicator was an anonymous individual whose acceptability might be determined by the stand taken by him on the issue under discussion. If he differed in position only slightly from *S* he would be regarded as fair and authoritative and bring about a shift in opinion, but if he differed greatly he would be regarded as incompetent and biased and fail to influence *S*'s opinion. To explore this factor and the one of involvement mentioned in the preceding paragraph, a factorially designed experiment will be required in which type of communicator (positive, negative, and neutral), type of issue (high involvement and low involvement), and distance between the communicator's stand and *S*'s own stand, are all systematically varied to permit analysis not only of the main effects but also of the more interesting interaction effects.

#### SUMMARY

The relationship between the attitude of the recipient and the position advocated in a communication was studied under conditions where a communicator not known to *S* presents

a point of view on a controversial issue which differs from that of *S* by varying amounts.

The topic discussed was prohibition of alcohol. The *Ss* came from a dry state where this was a lively issue. Each *S* indicated for a series of nine statements ranging from extreme dry to extreme wet: (a) the position most acceptable to him, (b) other acceptable positions, (c) the position most objectionable to him, and (d) other objectionable positions. Groups of *Ss* were selected whose stands on the issue were known (e.g., WCTU members) for comparison with unselected groups of college students. In all, 514 *Ss* were used. In a subsequent session "wet" groups received a "dry" or "moderately wet" communication; "dry" groups received a "wet" or "moderately wet" communication; unselected groups received a "dry," "wet," or "moderately wet" communication. After the communication, opinion measurements and reactions to communication were obtained. The findings were as follows:

1. When the distance between *S*'s own stand and the position advocated in communication is small, the communication is judged favorably, as fair and factual. With increasing distance, the favorable reaction is sharply reduced and the communication perceived as propagandistic and unfair.

2. The *Ss* whose own stands diverge widely from the position advocated perceive the communication as further removed from their own stand than it is ("contrast effect"). The present results indicate, though less clearly, that *Ss* whose own stands are close to the position advocated perceive the communication as closer to their own stand than is the case ("assimilation effect").

3. The most frequent result for *Ss* whose own stand diverges widely from that advocated in communication is to remain unchanged in their initial attitudes. More *Ss* with moderate positions closer to the stand in communication changed in the direction advocated.

It is suggested that the relative distance between *S*'s own attitude and communication along with *S*'s latitudes of acceptance and rejection for various stands on the issue may provide a basis for predicting reactions to communication and susceptibility to change. Further research varying degree of ego-involvement

ment in issues and attitudes toward the communicator is suggested.

## REFERENCES

1. FRENCH, J. R. P., JR. A formal theory of social power. *Psychol. Rev.*, 1956, **63**, 181-194.
2. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.* 1954, **49**, 325-329.
3. HEINTZ, R. The effect of remote anchoring points upon judgments of lifted weights. *J. exper. Psychol.*, 1950, **40**, 584-591.
4. HOVLAND, C. I. Effects of mass media of communication. In G. Lindzey (Ed.) Ch. 28, 1062-1103, *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954.
5. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.
6. HOVLAND, C. I., & PRITZKER, H. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257-261.
7. HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822-833.
8. KNOWER, F. H. Experimental studies of changes in attitudes: I. A study of the effect of oral argument on change of attitude. *J. soc. Psychol.*, 1935, 315-345.
9. LUBIN, A., & FISHER, S. Distance as a determinant of influence in a two-person continuous interaction situation. E.P.A. Talk. Title in *Amer. Psychologist*, 1956, **11**, 491.
10. MANSKE, A. J. The reflection of teachers' attitudes in the attitudes of their pupils. In G. Murphy, L. B. Murphy, & T. M. Newcomb (Eds.), *Experimental social psychology*. New York: Harper, 1937.
11. MURPHY, G., MURPHY, L. B., & NEWCOMB, T. M. *Experimental social psychology*. New York: Harper, 1937.
12. REMMERS, H. H. Propaganda in the schools: Do the effects last? *Publ. Opin. Quart.*, 1938, **2**, 197-210.
13. ROGERS, S. The anchoring of absolute judgments. *Arch. Psychol.*, 1941, No. 261.
14. RUSSELL, D. H., & ROBERTSON, I. V. Influencing attitudes toward minority groups in a junior high school. *School. Rev.*, 1947, **55**, 205-213.
15. SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48**, 135-141.
16. SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. (Mimeo. Rep.)
17. WEISS, W., & FINE, B. J. Opinion change as a function of some intrapersonal attributes of the communicatees. *J. abnorm. soc. Psychol.* 1955, **51**, 246-253.
18. WILKE, W. H. An experimental comparison of the speech, the radio, and the printed page as propaganda devices. *Arch. Psychol.*, 1934, No. 169.
19. WILLIAMS, A. C., & REMMERS, H. H. Persistence of attitudes concerning conservation issues. *J. exper. Educat.*, 1939, **40**, 354-361.
20. WILLIAMS, R. M., JR. *The reduction of intergroup tensions*. New York: Soc. Sci. Res. Council, Bull. 57, 1947.

Received November 30, 1956.



# COMPATRIOT AND FOREIGNER: A STUDY OF IMPRESSION FORMATION IN THREE COUNTRIES<sup>1</sup>

JEROME S. BRUNER

*Harvard University*

AND

HOWARD V. PERLMUTTER

*Massachusetts Institute of Technology*

A COMMONSENSE observation is the starting point of our inquiry. If one were to enter a room and be introduced to a Frenchman, a German, and an American, it would be likely that one's first impressions of the three would be such as to magnify the nationality difference between our new acquaintances. The Frenchman might appear more "typically French," the German more "typically German," and the American more "typically American." Our perceptual selectivity would register cues that differentiate the three in terms of their most evident difference: their nationality. To this commonsense observation we must add a second. It is also likely that, if one were an American, the Frenchman and German would seem more "typical" of their nationality than would the American to whom we had been introduced. And this would perhaps be the more so to the degree that we were well acquainted with Americans but had experienced little contact with Frenchmen or Germans.

The matter need not be limited to nationality. Introduced at the same time to two men, one as a professor and one as a businessman of our own nationality, likely as not we would seek the cues that permitted maximum differentiation of the two in terms of the difference of profession. And again, the degree to which this would be so might depend upon how well acquainted we were with practitioners of the two professions: the better acquainted, the less the likelihood of role stereotyping. If we were familiar with one of the professions, by

the same token, the other would undergo a stereotyping in our impression formation.

These commonsense observations; or rather, the hypotheses to which they give rise and with the testing of which this paper is concerned, lead to several thoughts about the nature of the process that has come to be called "the formation of first impressions." It suggests in the first place that under certain conditions, our impressions have the effect of accentuating the differences between people encountered briefly, and further, that there are other conditions that reduce this tendency to the accentuation of difference. Let us state the matter more formally in terms of a set of hypotheses.

*Hypothesis 1.* If a person be placed in a particular category on first encounter (nationality, role, occupation), the more widespread one's experience with diverse members of the category the less will this category membership affect the impression formed of the person encountered. Put concretely, if one knows Americans better than Frenchmen or Germans, the first impression formed of an American will be less determined by his nationality than in the other two cases. Or if one knows musicians better than theoretical physicists, the former occupation will affect a first impression less than the latter.

*Hypothesis 2.* If three pieces of wood, alike in all respects save their grain, be viewed simultaneously, an observer will be more likely to notice grain characteristics in the wood than if one of the pieces is presented alone. Put more formally, that attribute which distinguishes an array of objects will be most salient in viewing the array. Given a single object, there are no similar constraints on what an individual may notice (3). With reference to our own experiment, if the individual is asked to form impressions of an array of individuals differing only in nationality, then nationality will be a more determinative factor in impression forma-

<sup>1</sup> This research was jointly supported by the experimental project of the research program on international communication at the Center for International Studies, Massachusetts Institute of Technology, under the direction of Dr. Perlmutter, and ONR contract ori-0746 at Harvard University under the direction of Dr. Bruner. We are especially grateful to Dr. Renate Mayntz of UNESCO, and M. Michel Hurtig of the Centre National de la Recherche Scientifique of Paris for assistance in the collection and analysis of data.

tion than if the individual is set the task of forming an impression of a single individual.

With this much of an introduction, we may turn to the study designed to examine and test these hypotheses.

### DESIGN OF THE STUDY

The basic design of the study was simple in outline. By the use of a brief sketch, a stimulus person or stimulus persons were presented to our subjects (Ss). The Ss were then given the task of checking on a trait list those items they thought characteristic of the stimulus person as a "first impression." The stimulus persons presented varied in terms of their nationality (French, German, or American) and in terms of their occupation—unspecified, college professor, or businessman. Moreover, the impression formation situation varied, as required by our second hypothesis, in terms of whether the S had a single person of a particular nationality in mind when forming his impression or had in mind three people of different nationality: what we shall call the "single-impression" and the "triple-impression" situations. After filling in the impression trait list, Ss were asked to indicate, in a manner to be described, the bases upon which they had decided to check various traits. The Ss were all university students and themselves varied in nationality. There were university students from Boston; Paris and Dijon, in France; and a set of groups from Berlin, Hanover, and Cologne, in Germany. Students in each country, then, were forming an impression of a compatriot and/or of a foreigner. We may consider now the specifics of the design.

Thirty-seven groups of Ss were used in the "single-impression" situation, a third of them Germans, a third French, and a third American. The total sample was comprised of 576 Ss of whom 298 were male and 278 female. The groups who were given the test while assembled, varied in size from 13 to 29 Ss. Each S in these groups, as we have noted, formed an impression of a single stimulus person. The twenty-seven groups and the number of Ss in each is set forth in Table 1. Since no consistent sex differences were found, the data for both sexes were combined.

Only nine triple-impression groups were required to fulfill the design. The total population in these nine groups comprised 218 Ss

TABLE I  
THE NUMBER OF SUBJECTS IN EACH  
OF TWENTY-SEVEN SINGLE-IMPRESSION GROUPS

Subjects	Stimulus persons who are—								
	American			French			German		
	Prof.	Bus. man	Unspec.	Prof.	Bus. man	Unspec.	Prof.	Bus. man	Unspec.
American	19	13	21	15	18	22	16	18	19
French	29	22	18	29	28	27	28	28	27
German	20	20	20	20	20	19	20	20	20

TABLE 2  
THE NUMBER OF SUBJECTS IN EACH OF NINE  
TRIPLE-IMPRESSION GROUPS

Subjects	Stimulus persons who are—		
	Amer., Fr., & Ger. Profs.	Amer., Fr., & Ger. Bus. men	Amer., Fr., & Ger., Unspec.
American	26	28	30
French	26	25	23
German	20	20	20

of whom 108 were males and 110 females. Each S formed impressions of three stimulus persons of different nationality: in a third of the groups the stimulus persons were a French, American, and German professor; in another third, French, American, and German businessmen; and the remaining third, a French, American, and German person of unspecified occupation. The design and the number of Ss in each group is presented in Table 2. These Ss knew in advance, of course, that they were to form impressions of people of three different nationalities.

All Ss were given a booklet with the instructions not to look ahead in the questionnaire and to finish each page before turning to the next. Both single-impression groups and triple-impression groups were told: "You will be asked to form impressions of specific persons in this questionnaire." The latter were further instructed: "These people are from different countries."

The instructions and description of each stimulus person were as follows:

The object of this test is to determine the extent to which people are capable of judging a person from a few facts about that person. A brief characterization of a specific individual appears below. Read it care-



TABLE 3  
THE TRAIT LIST AND ITS GERMAN AND FRENCH EQUIVALENTS

Synonyms for "Energetic"					
Focused			Diffuse		
Trait	German	French	Trait	German	French
resolute	entschieden	résolu	bustling	geschäftig	agité
tenacious	zäh	tenace	animated	angeregt	animé
dogged	verbohrt	persévérant	spirited	munter	fougueux
determined	entschlossen	décidé	vivacious	lebhaft	vif
Synonyms for "Intelligent"					
Focused			Diffuse		
Trait	German	French	Trait	German	French
analytic	analytisch	analytique	wise	weise	sage
brilliant	geistreich	brillant	judicious	abwägend	judicieux
quick	schnell	alerte	level-headed	einsichtig	bien équilibré
logical	logisch	rationnel			
penetrating	durchdringend	perspicace	prudent	klug	prudent
astute	scharfsinnig	astucieux	sensible	vernünftig	sensé
clever	gescheit	adroit	sound	gesund	doué de bon sens
discerning	scharfblickend	doué de discernement	unaffected	ungeziert	simple
canny	schlau	avisé	sensitive	empfindsam	pourvu de flair
shrewd	listig	sagace	tactful	taktvoll	diplomate
cunning	verschlagen	ruse	understanding	verständnisvol	comprehensif
calculating	berechnend	calculant d'avance	unprejudiced	vorurteilsfrei	impartial
Synonyms for "Well-adjusted"					
Focused			Diffuse		
Trait	German	French	Trait	German	French
self-possessed	gelassen	maitre de lui	cheerful	fröhlich	joyeux
philosophical	philosophisch	philosophe	happy	glücklich	heureux
deep	tief	profond	satisfied	zufrieden	content de bon sort

fully and try to form an image of this person. You will be asked to record your impressions.

He is a very typical \_\_\_\_\_. There is general agreement among those who know him that he is intelligent, energetic, and well-adjusted. Now 42 years old, he is married and lives in a large city in \_\_\_\_\_."

For the American businessman, the *S* would be told:

"He is a very typical American businessman . . . who lives in a large city in the United States." For the French college professor, the subjects would be told: "He is a very typical French college professor . . . who lives in a large city in France," etc.

Next, *Ss* were asked to turn the page and choose from a list of 38 traits (see Table 3, which contains the trait names in English,

French, and German) those which "best characterized the person described." The *Ss* were allowed to choose as many as they wished.

After the trait list had been checked (in the case of the triple-impression group, after trait lists for all three stimulus persons had been checked), *Ss* were next asked the bases on which they had checked specific traits on the list. A series of letters were used to describe these *determinants* of each trait-choice. For each trait checked, *Ss* wrote:

E if the trait *energetic* of the stimulus person contributed to that choice,

I if the trait *intelligent* of the stimulus person contributed to that choice,

W if the trait *well-adjusted* of the stimulus person contributed to that choice,

S if the *social role* contributed to that choice,

N if the *nationality* of the stimulus person contributed to that choice.

Subjects could use *any* combination of the letters to describe completely the bases for each choice. They were encouraged to specify freely any other determinants of their choice, although few actually did.

Since we shall rely heavily upon the report of Ss concerning the determinants of their choice of traits, a word about the "meaning" of such reports is necessary. One cannot naively assume that Ss "know" what led them to check a particular trait. It is as conceivable that the degree of one's hostility toward a businessman-father is as much a determinant of choice as anything contained in the sketch. But it would be just as naive to assume out of hand that Ss, university students in this case, are completely incapable of "knowing" the basis for their selection of traits. The issue is not one that can be resolved. Yet, we have asked our Ss to indicate the basis of their choices. We take their responses as symptomatic not of the "true" basis of choice—whatever that may mean ontologically—but rather as a basis for inferring what underlies their choice. In fact, the only proper basis of inference would be further systematic variation in the trait lists used, in the instructions given Ss, and in the nature of the responses they gave by which we make our inferences about causes. We shall go ahead in this paper treating our Ss' reports on determinants *as if* they could be taken as a proper basis of inference. In a final section we shall reconsider the matter.

### *Design of the Trait List*

The traits used to characterize the stimulus persons were constructed in the following way, the activity being conducted in English: (a) A list of synonyms of the stimulus traits "energetic," "intelligent," and "well-adjusted" were gathered; (b) synonyms were divided into dichotomized groupings on the basis of certain dimensions. (See Table 3.) Synonyms of the stimulus trait "energetic," for example, were subdivided into "focused-energetic" and "diffuse-energetic." Under "focused-energetic" we put the traits: bustling, animated, spirited, and vivacious. These dichotomies were based on hunches concerning shifts in the meaning

of stimulus traits when connected with persons of various nationalities.

The final ordering of traits in the list was random. Recall that Ss were asked to report the determinants of each trait they checked. With the exception of two traits all traits on the list were found to be determined by the stimulus traits for which they were designed to be synonyms. So Ss checking "sensible" on the trait list would more often report that the "intelligent" characteristic of the stimulus person determined their choice rather than the fact that he was said to be "well-adjusted" or "energetic."

In both France and Germany the researchers were given the American trait list and were asked to translate this list into their respective languages. A number of problems arose. First, it was difficult to establish an exact lexical equivalent because either the language did not utilize the stimulus traits or the check list traits in the same way, or the lexical equivalent was not used with the same degree of frequency in each country according to the judgment of the respective researchers. In all cases, the list was translated by not less than three bilingual individuals native to the country, and a final discussion was held to decide which translation of the equivalent trait in the American trait list was to be used.

There is no practicable procedure for assuring that the trait lists in the three languages are either denotatively or connotively identical, and the results to be reported must be evaluated in the light of this fact. A close approximation to a test of identity is by the use of a matching procedure. We have asked English-German and English-French bilinguals to match items in the two lists to ascertain whether, within this universe of traits, identity matches could be obtained. But this does not obviate the contingency that the items matched across languages are not the *closest* synonyms possible. All that we can say is that we have taken precautions. For a fuller account of the problems of translating traits from one language into another, the reader is referred to the reports of Perlmutter, Mayntz, and Hurtig, (6) and Lenneberg and Roberts (4).

### RESULTS

One of our hypotheses was that Ss operating in the triple-impression situation will rely more heavily on nationality as a determinant than



will the single-impression *Ss*. A gross test of the hypothesis is provided by comparing the number of times on the average that *Ss* in the two situations indicated that their choice of a trait for characterizing the person was based upon knowledge of nationality. Grouping together all French, American, and German stimulus persons without regard to occupation and comparing the average number of times that *Ss* justified their choice of a trait by reference to nationality, we obtained the confirmatory results presented in Table 4.

Another way of stating these results is by reference to comparisons between specific single- and triple-impression groups. There are 27 single-impression groups as indicated in Table 1. For each, there corresponds a comparable record given under triple-impression conditions. In all 27 of the possible comparisons, the triple-impression group shows a greater reliance on nationality as a determinant of their responses, a result of a confidence level well beyond the .01 level as determined by the Dixon-Mood Sign Test.

Table 4 also confirms our other hypothesis that, in general, one uses nationality determinants less often in forming one's image of a compatriot than in forming an impression of a foreigner. This result was obtained in both single-impression and triple-impression situations. Comparisons may be made in this way. Take first the single-impression situation. For each group of *Ss*, let us compare the use of nationality as a determinant in forming an impression of a compatriot and of a foreigner. Thus, we will compare the use of nationality determinants for Americans judging Americans with their use of this determinant in judging Frenchmen and in judging Germans. We thus have two comparisons. We may also compare the use of nationality when Americans form an image of an American professor as compared with their usage in forming an impression of a German professor and a French professor. For *Ss* of all nationalities, 18 such comparisons are possible for the single-impression situation and another 18 for the triple-impression procedure. In 34 out of the 36 comparisons thus afforded, nationality is used more frequently as a determinant in forming an impression of a foreigner than of a compatriot, again a statistically highly reliable result.

Table 4 indicates, however, that under single-impression conditions, this tendency to

TABLE 4  
AVERAGE NUMBER OF TIMES THAT TRAIT IS  
CHECKED ON BASIS OF NATIONALITY  
OF STIMULUS PERSON

Group	French Subjects		German Subjects		American Subjects	
	Foreigner	Compatriot	Foreigner	Compatriot	Foreigner	Compatriot
Sing.-Impress.	4.6	2.8	3.9	1.4	3.3	1.3
Trip.-Impress.	6.8	6.8	6.6	4.5	5.2	4.3

use nationality more often on foreigners than on compatriots is considerably more marked than under triple-impression conditions.<sup>2</sup> The latter conditions seem to have the effect of making nationality salient even in sizing up a fellow national. We may note, for example, that with the French *Ss* operating under triple-impression conditions, nationality determinants are used as often on compatriots as on foreigners. It was in this subgroup that the two reversals in trend noted above were found.

We come now to several corollaries of the major hypothesis with which we have just been concerned. The first of these has to do with the specificity of the category into which a stimulus person is "coded." Recall that *Ss* were given sketches of a college professor, a businessman, and a person of unspecified vocation varying, of course, in nationality. Recall also that the *Ss* are themselves university students in working contact with professors perhaps more than with businessmen. It would follow then that the category "professors" would be more differentiated for them than the category "businessmen." We would hypothesize, then, that vocation would more often be used as a determinant in forming

<sup>2</sup> This tendency is statistically a highly reliable one. We test significance in the following way. Compare the difference of nationality determinants used in forming an image of a foreigner and of a compatriot under single- and triple-impression conditions. Eighteen such comparisons are possible. For example, for French *Ss* operating under single-impression conditions, we may compare the difference in number of nationality determinants used in forming an impression of a typical American professor and a typical French professor ( $3.00 - 2.10 = .90$ ) with the difference obtained for the same stimulus persons judged under triple-impression conditions ( $6.54 - 5.96 = .58$ ). In 17 of the 18 comparisons possible, the differences were greater for the single-impression groups, a result significant beyond the .01 level.

TABLE 5  
AVERAGE NUMBER OF TIMES THAT TRAIT IS  
CHECKED ON BASIS OF VOCATION OF  
STIMULUS PERSON

Stimulus Person	Sing.-impress. Groups			Trip.-impress. Groups		
	"Prof."	"Bus. man"	Bus. men Checked More Than Prof.	"Prof."	"Bus. man"	Bus. men Checked More Than Prof.
Amer. subjects:						
Amer.	2.1	6.0	+	4.2	3.9	—
Ger.	3.3	5.0	+	3.5	3.4	—
Fr.	3.3	5.8	+	2.8	3.3	+
Fr. subjects:						
Amer.	2.7	6.3	+	2.5	5.9	+
Ger.	4.5	5.9	+	2.9	4.7	+
Fr.	3.9	6.4	+	3.8	5.7	+
Ger. subjects:						
Amer.	3.1	5.0	+	3.1	4.7	+
Ger.	3.6	6.0	+	4.9	4.6	—
Fr.	3.9	4.6	+	2.6	4.4	+

an impression of a businessman than of a professor, at least amongst our university-based subjects. Such is indeed the case. Consider first the nine single-impression groups. For American Ss, for example, we can ask whether vocation is more often used as a basis for forming an impression of an "American businessman" or an "American professor." We can also compare German professor and businessman and French professor and businessman for this group of Ss and for Ss of other nationalities. This gives us nine possible comparisons. In all nine of these comparisons, the results come out as expected: vocation is more frequently named as a basis of checking list traits for businessman than for professor. These results are summarized in Table 5 where the material for the triple-impression condition is also presented. In the latter case, the picture is less clear: six of the nine comparisons are in the expected direction. If the 18 comparisons be grouped, the resultant figure of 15 in 18 is significant by the Sign test at the .01 level. If the reader examines the reversals in Table 5, he will note that they are more of the order of ties than of reversals in two of the three cases.

#### THE NATURE OF THE IMPRESSIONS FORMED

Our object in the present report is to consider principally the nature of the impression-forming process and the factors that may influ-

ence this process. In the preceding section, the emphasis has been upon determinants, and in the one that follows it will be upon the facets of the images formed by our Ss as inferred from the traits they actually checked in characterizing the various stimulus persons presented.

A full presentation of the traits ascribed to Americans, Frenchmen, and Germans of different occupation by respondents of these several nationalities would be a forbidding undertaking. There are 36 groups of Ss, each with a somewhat different kind of stimulus person, at least different in terms of conditions under which presented. And there are 38 traits which may be checked.

One approach is to choose a particularly interesting stimulus person—one whose image seems to loom large in the political and social sphere: the American businessman. We will limit our discussion to him as illustrative.<sup>3</sup>

To reduce the complexity of the image that emerges, we limit ourselves to a discussion of those traits that are checked by at least 50 per cent of respondents from a given country. These we may regard as "consensual" impressions. In Table 6 are set forth those traits in which such a consensus was found with respect to the image of the American businessman. All three nationalities, regardless of the method of testing, agree that the American businessman is "determined," and there is also rather wide agreement that he is "sensible." French respondents rather agree with their American fellows in seeing him as "satisfied" and "level headed"—rather a sobersided image—whereas the Germans appear to emphasize a constellation of "resolute," "tenacious," and "calculating." It is difficult, of course, to extract an organized image from a set of checked items on a trait list; and also from the items that go unchecked. For as Asch has so wisely pointed out and demonstrated (2), traits fuse together in an organization, and the organization is often dominated by a single trait. We have no way of determining from our data, for example, whether "calculating" is the organizer of many German impressions, or "resolute" of French images.

There is not a very sharp difference in the images formed by different nationalities, a fair and considerably above-chance agreement

<sup>3</sup> A detailed qualitative analysis of the German findings has been published in Germany (6). In this study, individual differences in the formation of impressions of foreigner and compatriot are discussed.



TABLE 6

TRAITS AGREED UPON BY HALF OR MORE OF SUBJECTS OF EACH NATIONALITY IN CHARACTERIZING AN AMERICAN BUSINESSMAN

Single-Impression			Triple-Impression		
Americans	French	German	Americans	French	German
cheerful	joyeux		cheerful	joyeux	
tactful	diplomate		tactful		
sound	doué de bon sens		satisfied	content de bon	
satisfied	content de bon			sort	
logical	sort				logisch
	rationnel	entschieden		résolu	entschieden
	résolu	zäh		tenace	zäh
	tenace	ungeziert	unaffected		ungeziert
	simple	geschäftig			geschäftig
bustling	sensé	vernünftig	sensible	décidé	entschlossen
sensible	décidé	entschlossen	determined	adroit	
determined			clever	heureux	naturlich
clever			happy		
happy			level-headed	bien équilibré	
natural					
level-headed	bien équilibré				
understanding					
	pourvu de flair				
	astucieux				
	maître de lui			maître de lui	gelassen
	avisé			avisé	
	calculant			calculant	berechnend
	d'avance			d'avance	
			shrewd		
				alerte	schnell
				animé	abwägend
					scharfblickend

being the rule for Americans, Germans, and Frenchmen checking a trait list for stimulus persons of any nationality. We shall not go into detail on this matter, for it is a subject with little issue. Partly it is an artifact of synonymy: an American who is "intelligent" and "well-adjusted" as presented in a sketch is more likely than not to be called "astute" by a fellow-American, "scharfsinnig" by a German, and "astucieux" by a Frenchman. But, then, so is a German stimulus person so characterized.

#### RECAPITULATION

Our first and perhaps most general hypothesis is that if objects that are alike in all respects save one are considered *together*, their difference in this one respect will be more critical in the impression one forms of the objects. Three identical triplets, differing only in the color of tie they are sporting, will be seen and interpreted more in terms of their tie-wearing habits than would be the case if each one were encountered singly and without the possibility of a simultaneous comparison. A corollary of this hypothesis to which we have addressed ourselves is: If in forming impressions of foreigners

and compatriots one is thinking in a comparative context, with the different nationalities in mind while forming one's impression, then the degree to which nationality will influence the impressions formed will be increased. The comparative context, often celebrated rather uncritically as a facilitator of panhuman perceptiveness, may not always have such an effect. In crosscultural contact, an individual may in fact exaggerate the impressions of differentness between foreigner and compatriot. Indeed, anthropologists have often been criticized for their description of the gulf between members of different cultures, particularly by those who, like Allport (1), have tried to emphasize the communalities in the human condition. It may indeed be the special comparative perspective of the anthropologist that produces this emphasis, if indeed it is such.

Our second hypothesis is similarly simple. If a person or object be "placed" or classified in an undifferentiated category with the members of which one has had little experience, the effect will be for the general properties of the undifferentiated category to have a greater

effect on the impression formed of the individual so placed than if the placement had been in a more differentiated category. Category membership, in brief, will have a more telling effect on the impression formed of one of its members in the degree to which the category is differentiated. More differentiated categories have less "stereotyping" effect than less differentiated ones. Thus, the nationality of a person will stereotype one's impression of him to the degree that the nationality is well or poorly known to one. So too with occupation. A college professor will do little stereotyping on the basis of being told that a man he is meeting teaches at a university. Should he know businessmen less well, the announcement that the man in question earns his living in commerce will have a greater effect on the impression formed. To put the matter more elegantly, a differentiated category is one in which there is a better representation of the likelihood with which specific traits are associated with all or most of its members.

Our two principal hypotheses leave much ground still unturned in the matter of how one forms impressions of compatriots and foreigners, but our findings do serve to sharpen up a few issues. Impression formation depends in massive degree on categorization processes. In essence, we place a person or thing in a category on the basis of a few minimal cues—like a statement of his nationality or occupation—and then proceed to "run off" along the lines of the higher probability attributes associated with people or events included in the category. As noted in a recent work of Bruner, Goodnow, and Austin (3), the attributes that will be emphasized under these circumstances are dictated at least in part by the nature of the discrimination that the impression-former has to make. If one should be asked for the characteristics of man that distinguish him from bears, one set of attributes will become salient. If the task is to distinguish man from all other species, we may have recourse to the old characterization of man as a featherless biped.

We turn at last to the methodological problem introduced earlier. We have asked our Ss first to check certain traits that characterize a person presented in a sketch, and then to indicate the basis for doing so—what in the sketch prompted the choice of a given trait. Much of our reasoning has been based on the latter of these data: notably, the reliance on

nationality as a basis for checking traits. It is fashionable to be apologetic about utilizing such direct reports as a basis for drawing inferences. We do not feel that the matter can be resolved within the compass of the data available to us. This much is clear. Highly consistent differences have been found, differences that have a congruency as one goes from comparison to comparison. Either our Ss *thought* that their choice of traits was being determined by certain features of the sketches presented to them or they were in some degree so determined. The proper skepticism requires that the matter be left open pending further research—research designed to compare direct reporting with such other indirect methods of response as one may find in the armamentarium of projective tests or psychophysiological procedures. Any one such procedure would be as suspect inferentially as the direct reports of our Ss, no more and no less so. Ideally, inferences should be drawn from a very wide sampling of different kinds of responses. Let the reader suspend judgment in the present case, or better still, let him consider whether at this stage of research it is better to commit a Type I or a Type II error.

One final word is in order. Impression formation is not a separate sort of cognitive activity and we are in the debt of Asch (2) for making this clear. It is a phenomenon that requires close analysis in terms of cognitive theory—whether the theory be associationistic, Gestalt, or the type of information-utilization theory that has informed the present enterprise.

#### REFERENCES

1. ALLPORT, G. W. *Becoming*. New Haven: Yale Univ. Press, 1955.
2. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
3. BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
4. LENNEBERG, E. H., & ROBERTS, J. M. The language of experience: a study in methodology. *Suppl. Int. J. Amer. Linguist.*, 1956, 22, No. 2 (Whole Mem. 13).
5. MAYNTZ, RENATE, & PERLMUTTER, H. V. Der Einfluss von Nationalitätsidentifikation und Berufsdefinition auf die Vorstellungsbildung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 1956, 8, 2.
6. PERLMUTTER, H. V., MAYNTZ, RENATE, & HURTI, M. A methodological note on the use of cross-national trait lists for research. Unpublished manuscript, Center for Int. Studies, M.I.T., Doc. -55/36, Nov. 25, 1955.

Received February 6, 1957. Prior publication.



## CASE REPORT

# THE TREATMENT OF TWO PHOBIC PATIENTS ON THE BASIS OF LEARNING PRINCIPLES<sup>1</sup>

VICTOR MEYER

*Institute of Psychiatry, University of London, Maudsley Hospital*

**T**HIS paper is presented as an illustration of the possibility of applying experimental method and some of the principles available in psychological theories of learning to the treatment of psychiatric symptoms. As such, it is somewhat similar to an earlier study by Jones (7). It differs from the latter, however, in that it demonstrates the importance of individual differences in personality structure and the manner in which such differences dictate modifications of therapeutic techniques.

The treatments to be described were based mainly on the principle of primary stimulus generalisation (5). According to this principle, a conditioned reaction is not only evoked by the original conditioned stimulus but also, to some degree, by a series of stimuli more or less similar to the original one. The more dissimilar the new stimulus, the less intense is the conditioned reaction.

Many phobic patients display anxiety so intense when presented with certain environmental situations that they are not amenable to any therapeutic approach. However, the generalisation principle suggests that the presentation of an environmental situation which resembles the original one but evokes a reduced amount of anxiety can be exploited therapeutically. In such a situation, one can attempt to substitute an adaptive normal reaction for the original unadaptive anxiety response. A graded continuum of similar situations can then be employed, moving towards the original one, so that the original unadaptive response can be eliminated. This is essentially one of the methods advocated by Jersild and Holmes (6) for the treatment of children's fear, and its theoretical basis has been elaborated by

such writers as Guthrie (4) and Wolpe (8). The latter, among others, has demonstrated the usefulness of these techniques in experiments on animals and has applied them successfully in the treatment of a series of psychiatric patients (9).

### SUMMARY OF CASE HISTORIES

For the purposes of this paper, a complete case history of each patient is not essential. A brief summary of the psychiatric notes will suffice.

#### *Case 1*

The patient, a married woman of 48 years, was admitted to the hospital for investigation of suspected temporal lobe epilepsy. Apart from blackouts, she complained of an excessive fear of going out.

No family history of psychiatric disorder or epilepsy was reported. The birth and developmental milestones seemed normal. She left school at the age of 15 and had a successful work history up to the time of her illness. Her marriage was fairly satisfactory, but there were no children. Before her illness, she was described as quick-tempered but a good mixer, cheerful, hardworking, and conscientious.

Her blackouts started at the age of 28 during pregnancy. Since then, she had suffered 18 blackouts. They all followed physical exertion or emotional upset and were usually preceded by an aura of dull sensation in the chest. She usually lost consciousness, fell limp, and hurt herself on several occasions. On recovery she vomited and felt "horrible."

At the age of 44 she began having peculiar "feelings." These consisted of epigastric sensations, followed by a "horrible" sensation of being about to fall. These episodes occurred frequently when she was about to go out or when she was outside on her own. Occasionally, when she was unable to control these feelings, she tended to panic and then to suffer one of her blackouts.

As a result of these symptoms, she developed a strong fear of going out on her own, of

<sup>1</sup>Thanks are due Dr. D. Hill, Dr. D. A. Pond, and Dr. J. D. Dewsbery, consultants to the Bethlem Royal and Maudsley Hospitals, for their permission, encouragement, and assistance in carrying out the treatment. A debt is also acknowledged to Dr. M. B. Shapiro, Mr. H. Gwynne Jones, and Mr. J. Inglis for offering valuable suggestions.

traveling on public conveyances and of having sexual intercourse for fear of an attack. From the age of 46 she refused to go out unaccompanied. Shortly before the admission to the hospital, she was occasionally unable to go out with friends or even with her husband. She gradually became more and more anxious, depressed (entertaining suicidal thoughts), and fearful. Losing confidence, she resigned from a responsible job and stopped caring for her home. She continued working until admission, but needed to be accompanied all the way to and from work. On admission she was pleasant, sociable, cooperative, rational, and well oriented.

The results of all investigations were negative with respect to temporal lobe epilepsy. Except for some mild abnormalities shown on the air-encephalogram, all the tests contraindicated organic involvement. In view of the possibility that the attacks might be of cardiac origin, she was referred to cardiologist. He could find no abnormalities in the cardiovascular system.

The patient was diagnosed as exhibiting a phobic state with blackouts of unknown nature and was referred to the Psychology Department for the symptomatic treatment of her phobias.

### Case 2

A man of 42 years was admitted to the hospital for investigation of blackouts. He also suffered from excessive tension and various fears, mainly related to entering enclosed and crowded spaces.

Two instances of psychiatric disorder were reported in the family history: His mother had a "nervous breakdown," and his elder brother experienced difficulty in going out alone. The patient's early development and childhood seemed normal. He left school at the age of 14 and worked successfully as a precision grinder. He had been happily married and had two children. Before his illness he was described as undemonstrative and timid, but affectionate as a father and husband.

At the age of 22 he had several fainting attacks on parades when serving in the armed forces. Following these, he developed episodes of apprehension which usually occurred in crowded places. He felt tense, flustered, and experienced an urge to run away in a panic.

Strange or crowded places and feelings of boredom were associated with the onset of these episodes. The presence of a friend and strong interest tended to prevent their development. Up to the age of 38, he could preserve control and avert panic. After that time, control became more tenuous. On four occasions, episodes of panic terminated in fainting attacks. As a result, he became worried about himself and sharply curtailed his social activities. Shortly before admission to the hospital, he had a feeling of apprehension in a barber's shop, panicked, mounted his bicycle and shortly afterwards collapsed. After this accident he refused to ride the bicycle and would not go out. He became anxious, tense, and depressed.

He was treated by his own physician with phenobarbitone with very little success. On examination for possible temporal lobe epilepsy, the results were negative. When admitted to the hospital, he appeared very tense and anxious, but rational, well oriented, and cooperative. He was reluctant to talk about himself and needed much encouragement.

Like the first case, he was referred to the Psychology Department for symptomatic treatment of a phobic state.

### TREATMENT

#### Case 1

The patient was interviewed twice with the aim of getting detailed information about her phobic symptoms. The rationale of the treatment was then explained and discussed with her. During this time the experimenter (*E*) endeavored to establish good rapport. According to the theory outlined, this is essential since *E* becomes part of the various environmental situations during treatment and, by establishing an effective relationship, he becomes a reassuring stimulus, tending to reduce anxiety.

No exact record of each treatment session was kept and, therefore, some of the description lacks precision. Since the most likely place for the abnormal behavior to occur was a door leading to the "outside" and the "outside" itself, the hospital roof garden, which has walls but no ceiling, was selected as a starting point for treatment. On the first occasion, the subject (*S*) entered the garden with *E*. She only reported mild "thumping in the stomach" throughout this session. Next day she reported



feeling perfectly well and volunteered to enter the garden on her own. When she had gone twice on her own without any signs of the symptoms recurring, the place of treatment was changed to the main hospital garden. She went out several times with *E* on successive days, then was required to go out on her own to meet *E*, and then to seek and find him. Since *S* appeared rather dependent on *E*'s presence, other staff members of the ward were asked to participate in the treatment without *S*'s knowledge of prearrangement by *E*. She went out with various nurses and patients for walks in the garden. Similar expeditions took place during the evening. Not more than two sessions (30 minutes each) were given each day. When asked to go into the garden on her own, she did without any difficulty. From that point, she was encouraged to take frequent walks into the garden during the daytime and the evenings.

The next stage consisted of *E*'s taking walks with the patient outside the hospital grounds. These started in back streets and were gradually extended into the main street; the distance covered increased gradually day by day. She also went out with other people. Eventually, she was able to set out on her own to meet *E* and others outside. She was taken out at least three times when feeling "upset." She took four bus rides accompanied by *E* and frequently volunteered to go shopping with other patients. After two weeks (16 sessions) of this treatment, she was able to make short expeditions on her own without difficulty.

She spent the subsequent 16 days going for walks with *E* and other people and on her own. Short trips on buses were included, and at least three times she went out on her own in the evening. She spent three weekends at home but was instructed to go out on her own only if she felt confident. She managed to take three short walks near her house. On the last day of treatment she traveled some three miles away from the hospital and back. During the whole treatment, she never reported any symptoms except occasional mild thumping in the stomach. Her behavior on the ward improved. According to the psychiatrist in charge and the nursing staff, she had very few "upsets" and was more cheerful. She felt confident and eager to be discharged.

The treatment took about five weeks, and

the patient claimed to have enjoyed it. When she was discharged, her relatives and friends were told that she should be encouraged to go out on her own provided she felt confident and not upset. It was planned that she should be seen once every two weeks in the outpatient department for follow-up purposes and, if necessary, to modify management.

*Follow-up.* For nearly five months the patient worked regularly and was unaccompanied when visiting the hospital. She took occasional solitary walks and was able to do her shopping. Each working day she was met at the station by a friend who traveled with her to the factory; on the return journey she was met by her husband. She managed to get to the station on her own, and on two occasions when the friend failed to meet her, she experienced no difficulties. A month after discharge from the hospital, she developed an acute chest pain for which she was treated by her doctor. Also, while on a bus with some friends, she had one of her blackouts without any warning and was unconscious for about 45 minutes. Despite these difficulties and worries, she continued to work and to go out on her own. At the end of the fourth month, she increased the frequency of her unaccompanied expeditions, and for the first time in two years attended adequately to her housework. She also went to work on her own seven times in succession when her friend was on vacation. She reported occasional feelings of anxiety, but she never panicked and was able to ward off the attacks by thinking about treatment and handling a mascot (a piece of candy) which she kept throughout the treatment. The last time she was seen, she felt cheerful, confident about going out on her own, and practically cured of "unreasonable fears."

Three days after the last interview, the patient died following one of her sudden blackouts when travelling with friends. The cause of her death was attributed to left ventricular failure; postmortem examination demonstrated heart disease.

### Case 2

Two extensive interviews with this patient indicated that his anxiety was not associated with any specific stimulus but was more generalised and might depend on a variety of

factors. For this reason, a modified technique was adopted, aimed at enabling *S*, by a process of conditioning, to cope efficiently with his anxiety no matter where it arose. First, a course of systematic desensitisation was given, based on Wolpe's (9) relaxation techniques. This form of treatment failed, since the patient reported that he could not evoke any increased anxiety by thinking about stressful situations. He remained tense and anxious throughout the sessions, but reported no changes in feeling; GSRs similarly indicated no disturbances.

The conditioning of anxiety-relief responses was then attempted by means of a technique similar to that described by Wolpe (9). An inductorium with a maximum inflow of 6 volts was applied to two fingers; the strength of the current was controlled by turning a knob. The patient was told to say aloud "calm yourself" when the shock became unbearable. As soon as the patient said this, *E* switched off the current. Wolpe has reported that many of his patients experienced relief from anxiety in disturbing situations when they used these words associated with the release from shock-induced tension.

Since it appeared that *S* was most likely to manifest his symptoms in a crowded cinema, it was decided to initiate a procedure similar to that used with the first patient. It was planned to visit the local cinema at a time when it is usually relatively empty and then gradually delay visits, going later and later each day until eventually the visit would be made at a time when the cinema is commonly quite full. The effect of boredom was also taken into account since the patient saw at least part of the same picture each day. During the first week, the patient received 5 or 6 electrical conditioning trials immediately before visiting the cinema with *E*. After some thirty to forty minutes, both *S* and *E* returned to the hospital, where 5 or 6 further conditioning trials were given.

After seven days of this treatment, there was no change in *S*'s behavior. During most of the visits to the cinema, he reported feelings of tension and an urge to leave. Following a weekend at home, he said that he still felt depressed, anxious, and tense, and although the "calm yourself" technique helped him slightly, he considered that he was generally unimproved by the treatment.

In every respect, the patient's reactions to the treatment differed from those manifested by Case 1. Whereas Case 1 reacted very quickly to the treatment and a good rapport was easily established, Case 2 remained aloof and detached.

At this stage, Case 2 was reconsidered. A plausible hypothesis was derived from Franks's study (2) concerning the conditionability of extraverts and introverts and from Eysenck's theory of anxiety and hysteria (1). From these ideas, it seemed to follow that Case 1 would condition better and extinguish less rapidly than Case 2. Similarly, both cases should be more neurotic than average, and Case 1 should be more introverted than Case 2. These inferences were supported by the patients' scores on Franks's eyeblink conditionability measure and on the Maudsley Personality Inventory. On the theory<sup>2</sup> that inhibitory drugs decrease conditionability and heighten satiation effects, Case 2 was given 10 milligrams of dexedrine (excitant) for four days. As soon as the effects of the drug became apparent, he was submitted each day to the same treatment procedure as during the previous week.

Throughout this second stage of treatment, the patient felt cheerful and relaxed and enjoyed going to the cinema. At no time did he display any symptoms. On the second day, he was left alone in the cinema. On the last day, he managed to stay alone in a crowded cinema without any difficulty. During the following weekend he went home but took 10 milligrams of dexedrine each day. He traveled alone on buses, rode his bicycle, and went to a cinema with his wife. He reported no feelings of disturbance. On his return from the weekend, the dosage of dexedrine was reduced by 2.5 milligrams a day with an equivalent amount of placebo substituted without the patient's knowledge. The conditioning procedure and the visits to the cinema on his own continued daily after the administration of the drug and the placebo. On the fourth day, he received 10 milligrams of placebo only. During this stage

<sup>2</sup> There is some evidence to support this theoretical position. Unpublished studies at Maudsley Hospital on amobarbital sodium and dexamphetamine sulfate by Franks and Trouton, and on arecoline and methyl atropine by Franks, Laverly, and Trouton have yielded results consistent with the notions applied therapeutically in the present study.



of treatment, he gradually became tense and anxious and reported feeling slightly depressed. After each treatment session, however, he said that he felt better temporarily and at no time did he feel tense and anxious in the cinema.

During the next weekend at home, he was given one placebo tablet for the Saturday and none for the Sunday. He coped with traveling and riding his bicycle, but felt somewhat tense and anxious. He maintained that the cause of these symptoms was the fear that the treatment might fail and that he might have to stay in the hospital. He expressed a strong desire to leave, saying that he had not felt this desire before the treatment started. He also claimed that his present anxiety and tension had a "different quality since he could find a reason for these feelings."

The first day after his return from this weekend, he was given a final test. He had to travel on a bus some three miles away from the hospital, return, and then visit, unaccompanied, another very crowded cinema. He reported that before the bus ride he felt tense and worried, but on his return, he had recovered, regained his confidence, and felt cheerful. He also enjoyed the cinema and reported that he "never felt better."

The rationale of the treatment was explained to him and every step of the treatment (except the substitution of placebo) was discussed with him. After the final test he was discharged.

*Follow-up.* It was planned to see the patient once every two weeks. He kept his first appointment only, reporting considerable improvement in travelling and in visiting crowded places. He had also gained confidence in himself and in relation to strangers and his superiors. He still occasionally got his "queer feelings" of worry, strain, and fear of fainting, but these were very mild and not very disturbing. The only difficulty he experienced was when riding his bicycle, but he continued to cycle to and from work. Two weeks later he sent a letter, saying that he could not keep the appointment since he was in financial difficulties. Apart from these difficulties, he was managing very well and had mastered his fear of cycling completely. He could cope easily with any uncomfortable feelings he experienced.

Three months later he sent another letter,

informing *E* that he was progressing very well. He had taken his children to the coast for a holiday, cycled to and from work without any difficulty, and visited cinemas regularly.

#### DISCUSSION

This paper attempts to demonstrate the application of some aspects of learning and the personality theory to the symptomatic treatment of psychiatric patients. At present, the experimental findings are not completely consistent, and the theories themselves are in early stages of development. The available findings, however, suggest rational experimental techniques for symptomatic treatment.

It cannot be strongly maintained that a mere stay in the hospital could account for the recovery from the symptoms, since both patients stayed in hospital well over a month prior to treatment, and according to their verbal account and the observation of their behavior in stressful situations, there was no indication of any improvement. In the present state of knowledge, however, it cannot be argued with any degree of assurance that the improvement of these patients' long-standing phobic symptoms was due to the treatment given. Even if there was sufficient evidence to maintain that the treatment was responsible for the improvement, so many relevant factors were uncontrolled that one would not know which aspect of the treatment was relevant and which irrelevant. These relationships can only be established by more extensive and well-controlled research.

Moreover, the usual objections made against symptomatic treatment on a priori grounds lack force. It seems plausible to argue, as Jones (7) has, that "Much evidence points to the fact that neuroticism is largely a constitutional defect for which no effective radical therapy is yet available. The individual of neurotic constitution in certain environmental circumstances develops certain symptoms. The rational therapeutic approach is then to treat the symptoms and to modify the environment so as to avoid their recrudescence or the development of fresh symptoms. The more specific the treatment, the more likely may be its success."

Several implications of theoretical and practical interest emerge from this study: (a) Since both cases manifested blackouts of an

unknown nature, particularly Case 1, the phobic symptoms presented were not certainly known to be functional in origin. The treatment, however, was oriented towards a functional analysis of the disorder. (b) It would be desirable to know whether, following a course of treatment like that described, the neurotic reactions are eliminated or merely overshadowed by a stronger normal reaction. Although Wolpe (8) provides some evidence that neurotic reactions in cats can be eliminated, the question must remain open on the present evidence. (c) Although on clinical grounds Case 1 was regarded as an hysterical personality and Case 2 as a dysthymic, objective psychological tests indicated the reverse. It seems likely, moreover, that the etiology of the phobic states must have been quite different in the two cases. One would like to know how two patients showing such a considerable difference in conditionability came to develop more or less similar patterns of abnormal behavior. (d) The results obtained during the course of treatment are consistent with Eysenck's theory (1) concerning the dynamics of anxiety and hysteria. They also indicate, in accordance with Eysenck's view, that for the purpose of treatment, adequate regard must be paid to the importance of certain individual differences as related to the disorders presented. (e) The use of drugs may facilitate and speed up successful treatment of this type, and this theory of individual differences may give a rational basis for the selection of drugs.

#### SUMMARY

The treatment of two phobic patients has been described. Case 1 manifested an excessive fear of going out on her own; Case 2 displayed disabling symptoms in the form of an excessive fear of going into enclosed and crowded places. Both cases had blackouts of unknown nature.

The treatment program for both cases was mainly based on the principle of primary

stimulus generalisation. Owing to the more "diffuse" nature of the anxiety displayed by Case 2, an additional simple conditioning technique was employed.

Case 1 responded to the treatment immediately and an improvement in her general behavior was observed. The effects of treatment persisted for nearly five months, when the patient died of left ventricular failure.

Case 2 failed initially to respond to treatment. An attempt was made to account for this failure in terms of Eysenck's theory of anxiety and hysteria. From the theory, inferences were made with respect to the patient's behavior on objective psychological tests; the results were consistent with the inferences. The treatment was modified according to the theory and the patient responded as expected. Four months after discharge from the hospital, the patient seemed to be managing well and to feel much improved.

#### REFERENCES

1. EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, 101, 28-51.
2. FRANKS, C. M., & LAVERLEY, S. G. Sodium amytal and eyelid conditioning. *J. ment. Sci.*, 1955, 101, 654-663.
3. FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, 52, 143-150.
4. GUTHRIE, E. R. *The psychology of learning*. New York: Harper, 1935.
5. HULL, C. L. *Principles of behavior*. New York: Appleton, 1943.
6. JERSILD, A. T., & HOLMES, F. B. Methods of overcoming children's fears. *J. Psychol.*, 1935, 1, 75-104.
7. JONES, H. G. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.*, 1956, 52, 414-419.
8. WOLPE, J. Experimental neurosis as learned behavior. *Brit. J. Psychol.*, 1952, 43, 243-268.
9. WOLPE, J. Reciprocal inhibition as the main basis of psychotherapeutic effects. *Arch. Neurol. Psychiat.*, 1954, 72, 205-226.

Received October 26, 1956.



## CRITIQUE AND NOTES

### PERCEPTION OF DISTURBING AND NEUTRAL WORDS THROUGH THE AUTOKINETIC WORD TECHNIQUE<sup>1, 2</sup>

SARNOFF A. MEDNICK, ALAN HARWOOD<sup>3</sup>, AND JACK WERTHEIM

*Harvard University*

THE Autokinetic Word Technique (AWT) presents a situation in which the subject (*S*) is placed in a completely darkened room and asked to read and report words and sentences being written by an apparently moving but actually motionless point of light. A previous study (1) indicates that *Ss* do indeed report the perception of words and sentences. The use of the term "perception" is perhaps debatable, but the authors feel justified in its use by the similarity of present operations to those involved in the study of such illusions as the Necker cube.

One of the important uses to which the AWT may appropriately be put is the investigation of nonautochthonous perceptual variables. In this study, it was hypothesized that in the AWT situation "disturbing" words would be perceived with greater latency and less frequency than "neutral" words.

#### METHOD

A word association test (WAT) of 40 words was administered to 12 male and 12 female paid volunteer college students. All of the words used were drawn from the recent Minnesota standardization of the Kent-Rosanoff Word Association Test on a college population (2). They all have a Thorndike-Lorge AA rating (3), indicating that they are commonly used in the English language. On the basis of probability of response (2), five disturbing and five neutral words were picked for each *S* from his WAT results. In cases where, due to ties, a decision could not be made solely on the basis of probability, the latency of the response served as the deciding factor. The median probability and latency of responses to the disturbing words were .001 and 2.10 sec., respectively; the median probability and latency of responses to the neutral words were .622 and 1.30 sec., respectively.

The *Ss* were then introduced to the AWT situation and instructed that we were going to write sentences with the point of light. (These *Ss* had had two hours of previous experience in the AWT situation in another study.) Before each sentence was written they were informed that they would be given one of the words in

the forthcoming sentence. This word might appear at the beginning, middle, or end of the sentence. The *S* was to inform *E* when he saw this word.

Ten trials were then given in which the stationary pinpoint of light was turned on for 90 seconds along with a cheap electric motor that produced a convincing, variable noise. On five of the trials, neutral words from the WAT were the suggested words. Disturbing words from the WAT were the suggested words on the other five trials. Whenever *S* reported seeing the suggested word during the 90-second period, his response latency was noted. When the light was turned off, *S* was asked to report the complete sentence that had apparently been written by the point of light. *E* was present in the blacked-out room with *S*, who was seated approximately nine feet from the point of light.

#### RESULTS

The two major measures of this study were (a) whether or not *S* reported perceiving a suggested word, (b) the latency of his reported perception. Of the 10 suggested words, *Ss* reported failing to see a mean of 3.73 words. *Ss* failed to see a mean of 2.09 of the five disturbing words and a mean of 1.64 of the five neutral words. This difference, evaluated by the paired replicates test (4), was significant at the .01 level. Thus, *S* failed to report seeing more of the disturbing than neutral words.

The mean latency of report of perception of the suggested words was 58.6 seconds. The perceived disturbing words had a mean latency of 64.4 seconds. The perceived neutral words had a mean latency of 52.7 seconds. The difference between these values was significant at the .01 level by the paired replicates test. Postponement of the report of perception of the disturbing word could be the result either of a simple unfilled delay of response or of the perception of many words preceding the disturbing word. The mean number of words preceding the suggested word was 1.02 words. The disturbing words were preceded by a mean of 1.08 words; the neutral words were preceded by a mean of .96 words, a difference that is not significant. The delay of report of perception of disturbing words cannot therefore be attributed to the perception of other intervening words.

The mean number of words reported in the sentences built around the disturbing and neutral words were 2.54 and 2.64, respectively. This difference, evaluated by the paired replicates test, was not significant.

<sup>1</sup> This research was supported by the Laboratory of Social Relations, Harvard University, and the Social Science Research Council.

<sup>2</sup> Presented to the 1956 meeting of the American Psychological Association.

<sup>3</sup> Mr. Harwood was the recipient of an SSRC Undergraduate Research Stipend, 1956, for work on this study.

## SUMMARY AND CONCLUSIONS

Disturbing and neutral words were selected individually for Ss by means of a word association test. It was then suggested to these Ss that an apparently moving but actually motionless point of light in a completely darkened room (autokinetic effect) would write a sentence containing a given word. The given word was either a disturbing or neutral word.

The Ss reported perceiving fewer of the disturbing words. Those disturbing words that were perceived had a greater latency than perceived neutral words.

## REFERENCES

1. RECHTSCHAFFEN, A., & MEDNICK, S. A. The autokinetic word technique. *J. abnorm. soc. Psychol.*, 1955, **51**, 346.
2. RUSSELL, W. A., & JENKINS, J. J. *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test*. Office of Naval Res. Tech. Report No. 11, 1954, Contract No. N8.
3. THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words*. New York: Teachers College, 1944.
4. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: Amer. Cyanamid Co., 1949.

Received October 1, 1956.

## LEARNING THEORY AND "OPPOSITE SPEECH"

ARTHUR W. STAATS

Arizona State College at Tempe

**L**AFFAL, Lenkoski, and Ameen (4), have recently reported the case of a schizophrenic patient demonstrating what they call an "opposite speech" syndrome. The authors state that the syndrome is basically a reversal in certain language usages, especially in the use of "yes" and "no." Feeling that the speech reversal is an expression of repressed hostility, they suggest that opposite speech copes with the hostile impulses by disrupting communication and rejecting other people or by allowing the "verbalization of ideas which the patient consciously rejects" (4, p. 412).

Laffal *et al.* conclude that learning theory has little to offer in this instance or others of pathological speech. "... [T]he so-called pragmatic and interpersonal functions of language behavior have been largely neglected by learning theorists in favor of the meaning or semantic aspects" (4, p. 412).

However, a number of recent experimental and theoretical studies (such as 1, 2, 3, 5, 6, 7, 9, 10) have been very much concerned with similar problems of language behavior. Several points can thus be made from a learning theory approach which have relevance for "opposite speech." Skinner, for example, generally defines verbal behavior as "behavior which is reinforced through the mediation of another organism" (7, p. 20). Instead of asking, as do Laffal *et al.*, "what uses [needs] the opposite speech may serve" (4, p. 412), reinforcement theory is concerned with the conditions which increase the probability that the response will occur again in similar stimulus situations or which strengthen the response.

It has been shown experimentally (1, 3, 9, 10) that social reinforcement which follows verbal responses has the effect of increasing their frequency. These studies have used various social reinforcements. The question is thus raised regarding the reinforcements involved in opposite speech. One possibility may be taken from a discussion of generalized reinforcers by Skinner.

The child who misbehaves "just to get attention" is familiar. The attention of people is reinforcing because it is a necessary condition for other reinforcements from them.... The attention of someone who is particularly likely to supply reinforcement—a parent, a teacher, or a loved one—is an especially good generalized reinforcer and sets up especially strong attention-getting behavior (8, p. 78).

It is suggested that attention, operating as a generalized reinforcer, could account in part for the origin and maintenance of opposite speech. The interview quoted by Laffal (4, p. 411) is a good demonstration of how the patient's opposite speech could elicit attention from others. Other reinforcers would also be effective in strengthening reversed language. For example, the patient's reversals were said to "not extend to... thought or action. Thus, if asked if he wishes a cigaret, the patient may say, 'No' instead of 'Yes, I do,' but he accepts and smokes the cigaret" (4, p. 412). Giving the cigaret to the "No" response is a reinforcement and would be expected to strengthen that response, i.e., raise the probability that the patient would again say "No" in the same situation. Reinforcements such as these, without con-



comitant admonition, would be unlikely to occur in the normal person's social environment.

Certain implications are derivable from this interpretation. If the opposite speech is maintained by positive reinforcement, then lack of such reinforcement should lead to extinction of such behavior. For example, withholding the cigaret should weaken the strength of opposite speech, and giving the cigaret to correct speech should strengthen that type of response.

Another example of learning theory formulations applicable to opposite speech is offered by Dollard and Miller (2). They discuss repression in terms of reinforcement theory. Certain thoughts arouse anxiety. Cessation of thinking those thoughts reduces anxiety. Thus, stopping thinking about that topic becomes a well-learned response. The same analysis can be applied to speech. Dollard and Miller (2) give an example of a group of people who change the topic of a conversation because it arouses anxiety, and state, "people tend to learn to avoid unpleasant topics of conversation" (2, p. 199).

It could be said that the schizophrenic patient's verbal behavior, when it is not confused by reversal, elicits an anxiety response in him, perhaps because of its typical content. Confused ways of speaking, and perhaps even of thinking, would therefore be anxiety reducing. This rationale could be extended to obsessive thinking in addition to other types of confused schizophrenic speech and thought. In addition, for the schizophrenic, communication with others which is understandable probably introduces touchy subjects which arouse anxiety. Reversed verbal behavior and other confused speech may reduce anxiety when it produces breakdown of the communication and cessation of the anxiety producing subject matter or of the bothersome conversation itself.

Why the schizophrenic's speech might elicit anxiety in him need not herein be elaborated, since the speech symptom is the relevant topic. It could be stated, however, that the unhappy life situation of an adult schizophrenic probably elicits thought and speech which are not positive secondary reinforcers, but instead arouse anxiety. It is also probable that the lack of success of the schizophrenic's

life behavior evokes verbal behavior from others which is anxiety producing for the schizophrenic. Verbal and nonverbal behavior of the schizophrenic which would avoid this anxiety would thus be well learned.

In concluding, it should be stated that this note is not intended as a complete analysis of the opposite speech of the schizophrenic patient. Perhaps it points out that hypotheses which apply to pathological language can be derived from a learning theory approach. At any rate, it is suggested that learning theory has reached a state where it has something to offer clinical theory and practice. The complex activities to which contemporary learning theory addresses itself, and not without success, indicate that the approach can no longer be ignored.

#### REFERENCES

1. COHEN, D. J., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106-110.
2. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
3. GREENSPOON, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctor's dissertation, Univer. of Indiana, 1950.
4. LAFFEL, J., LENKOSKI, L. D., & AMEEN, L. "Opposite speech" in a schizophrenic patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 409-413.
5. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
6. MOWRER, O. H. The psychologist looks at language. *Amer. Psychologist*, 1954, **9**, 660-694.
7. SKINNER, B. F. *Verbal behavior*. William James Lectures, Harvard Univer., 1947 (mimeo.).
8. SKINNER, B. F. *Science and human behavior*. New York: MacMillan, 1953.
9. TALBOT, E. The effect of note taking upon verbal responses and its implications for the interview. Unpublished doctor's dissertation, Univer. of California at Los Angeles, 1954.
10. VERPLANK, W. S. The control of the content of conversation: reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668-676.

Received October 22, 1956.

# THE INFLUENCE OF ETHNIC ATTITUDES ON REASONING ABOUT ETHNIC GROUPS<sup>1</sup>

NOR IAN M. PRENTICE<sup>2</sup>

Harvard University

**I**N THE abundant literature on ethnic prejudice, studies exploring the effects of tolerant and intolerant attitudes on reasoning about ethnic groups are rare. In one of the few relevant investigations, Thistlethwaite (5) demonstrated that prejudiced subjects biased their reasoning against ethnic groups. Moreover, an account by Allport (2) of unpublished research by S. B. Sells suggested that tolerant as well as intolerant subjects may bias their reasoning about ethnic groups in directions resonant with their attitudes, a finding in opposition to the traditional characterization of tolerant persons as rational and objective in their appraisal of ethnic groups. The account was nonetheless consistent with the evidence from related attitude areas that opposing social attitudes conduce to distorted reasoning in directions congruent with these attitudes (4). The aim of the present experiment was to examine more rigorously the influence of ethnic attitudes on reasoning logically about ethnic groups. The prediction was made that *tolerant individuals would show bias in favor of ethnic groups while intolerant individuals would show bias against ethnic groups on a test of syllogistic reasoning involving ethnic content.*

## METHOD

### Test Measures

1. *Attitudes.* Ethnic attitudes were measured by an 18-item form of the California Ethnocentrism (E) Scale, composed of an equal number of statements disparaging Jews and Negroes.<sup>3</sup> Authoritarian attitudes were appraised by a 12-item form of the California Fascism (F) Scale.<sup>4</sup>

<sup>1</sup> This report is based on portions of a doctoral dissertation completed at Harvard University in the Department of Social Relations in 1956. The author wishes to express particular indebtedness to his principal advisers: Drs. G. W. Allport, G. Lindzey, and R. W. White. Drs. N. Kogan and D. J. Levinson generously read this report and offered many valuable suggestions.

<sup>2</sup> Now at Judge Baker Guidance Center.

<sup>3</sup> Items selected from Adorno *et al.* (1) were numbers: 8, 22 (pp. 68-69); 13 (p. 70); 2, 31 (pp. 110-111); 37 (p. 117); and 1-6 of Subscales A and B (p. 142).

<sup>4</sup> The form given was taken from the FERPT Battery used by Dr. D. J. Levinson at Harvard University, Department of Social Relations, 1952. With one exception ("More than anything else, it is good hard work that makes life worthwhile"), the items are either identical or slightly rephrased from Adorno *et al.* (1): b-44 (p. 248); b-21, b-42, c-13, c-25, c-34, c-37 (p. 255); a-9, e-4, f-26 (p. 256); h-18 (p. 257).

2. *Objective reasoning about ethnic groups.* The ability to reason logically about ethnic groups was provided by performance on a syllogism test constructed by the author called, "The Woodbridge Reasoning Test."<sup>5</sup> This test consisted of 48 syllogisms randomly arranged which the subjects (Ss) rated for logical validity. The Ss were instructed to rate the conclusions to the syllogisms as following logically or not necessarily following logically from the premises. Of the 48 syllogisms, 24 were logically valid and 24 logically invalid by this definition. Each group of 24 contained 8 subgroups of three syllogisms. The three syllogisms in each of these groups were alike in validity, syllogistic mood, approximate sentence length, and linguistic style, but unlike in content. The premises and conclusions of one syllogism were tolerant, those of a second were intolerant, while those of a third were neutral; that is, designed not to produce any special affective reaction. The content of none of the affective syllogisms was formally identical to that in the E scale, though both measures dealt with Jews and Negroes exclusively. In contrast with the syllogisms, all of the E scale items were negative; i.e., critical of Jews and Negroes. Content for some affective syllogisms was adapted from Thistlethwaite (5), while that for others was drawn from stereotypes of minority groups enumerated by Allport (2).<sup>6</sup> The following triad in which all the conclusions are logically invalid illustrates the construction:

### Tolerant

If no one who is especially loud, noisy, and overbearing is admitted to a country club;  
And if many Jews are admitted to a country club;  
Then: No Jews are especially loud, noisy, and overbearing.

### Intolerant

If no one who mixes especially well with other people is aloof from the community activities;  
And if many Jews are aloof from the community activities;  
Then: No Jews mix especially well with other people.

<sup>5</sup> This test, together with a table indicating syllogism construction, has been deposited with the American Documentation Institute. Order Document No. 5297 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting \$1.25 for 35 mm. microfilm or \$1.25 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

<sup>6</sup> Dr. S. B. Sells graciously offered the syllogism test devised by him previously. Although none of his syllogisms were used in their original form, they provided structural models for the test devised herein. Dr. M. H. Thompson, now of the Philosophy Department, University of Chicago, kindly assessed the correctness of the syllogisms in their final form.



## Neutral

If no one who is especially interested in Egyptian art is concerned with Mexican art;  
And if many dealers are concerned with Mexican art;  
Then: No dealers are especially interested in Egyptian art.

Approximately 35 minutes were allowed for completion of the syllogism test.

An error score formula was derived on the assumption that the influence of ethnic attitudes would lead to reasoning errors in predictable directions on syllogisms with particular ethnic content. Thus, tolerant individuals were expected to have more errors on reasoning items that were valid but contained intolerant content and on reasoning items that were invalid but contained tolerant content. These same individuals were expected to have fewer errors on reasoning items that were valid but contained tolerant content and on reasoning items that were invalid but contained intolerant content. The converse was predicted for intolerant individuals. Thus errors in judgment of validity were expected to be systematically rather than randomly distributed across the reasoning items with affective content. The error score used for the analysis was as follows:

$$\text{Error Score} = [(AV - PV) + (PI - AI)]$$

where

- AV = Anti-Valid: number of errors on valid syllogisms with intolerant content.  
PV = Pro-Valid: number of errors on valid syllogisms with tolerant content.  
PI = Pro-Invalid: number of errors on invalid syllogisms with tolerant content.  
AI = Anti-Invalid: number of errors on invalid syllogisms with intolerant content.

The use of a difference score in the formula controlled for the level of reasoning ability. Neutral items were included in the syllogism test to provide variety of content and for other purposes, but scores on these items were not used in the analysis herein.

## Subjects

The tests were taken anonymously by 128 white Christian male and female freshmen enrolled at the Boston University School of General Education in the early Fall of 1955 who volunteered for a "study of social attitudes."<sup>7</sup> The Woodbridge Reasoning Test, E and F Scales, were administered in that order. None of the Ss had had courses in psychology or formal logic.

## RESULTS

Product-moment correlations between the E and F scores and the mean error scores on the Woodbridge Reasoning Test were  $-.22$  and  $-.30$ .

<sup>7</sup> The cooperation of Dr. G. N. Eddy and Mr. R. Smart, Social Science Department, Boston University College of General Education, in making Ss available for this study and prior pilot studies, is gratefully acknowledged.

TABLE 1  
COMPARISONS OF MEAN ERROR SCORES WITH  
HYPOTHETICAL ZERO MEAN SCORES ON  
WOODBIDGE REASONING TEST FOR  
INDIVIDUALS HIGH AND LOW  
ON E AND F SCALES

	N	$\bar{X}$	$t$	$p$
Low E	31	1.74	4.25	<.001
High E	32	.63	1.27	Reversal
Low F	30	1.57	3.75	<.001
High F	29	-.41	-.88	<.19

Note.—Values are for one-tailed  $t$  test following Lord as cited by Mosteller and Bush (3).

These values are significant beyond the .02 and .01 levels (one-tailed test), respectively. The correlation coefficients are negative since stronger intolerant attitudes (high E and F scores) are associated with less reasoning bias in the tolerant direction.

A subsequent analysis compared the mean error scores on the Woodbridge Reasoning Test made by individuals in the upper and lower quarters on the E and F scales, with hypothetical mean error scores of zero (unbiased reasoning). As the error score formula is applied, positive error means will be associated with the lower quarters of the E and F scales and negative error means with the upper quarters if respondents are biasing their reasoning in a direction congruent with their attitudes. Table 1 presents the results from this analysis. As indicated, in both the lower E and F quarters, significant bias operating *in favor* of ethnic groups was demonstrated (<.001 in both instances). However, in the high E and F quarters, significant bias operating *against* ethnic groups was absent, although the upper quarter F mean tended in the predicted direction. The absence of intolerant bias can be presumptively related to the composition of the sample. While the neutral item mean for both the attitude scales is 4.00, the item mean for the E scale was 2.72. Thus, as a whole, the sample with respect to the E scale was composed predominantly of tolerant individuals. The sample with respect to the F scale was somewhat more normally distributed with an item mean of 3.80. In this connection, an important distinction between relative and absolute position along these attitude continua should be made. In the present study, for example, the upper quarter E group, although extremely intolerant relatively, tended to be slightly tolerant in terms of the total possible range of ethnic attitudes measured by the scale. Indeed, as indicated in Table 1, the High E group showed a tendency to bias in the tolerant direction. Clearly, in considering extremes, it is important to be concerned not only with the samples at hand but with the theoretical universe as well.

## DISCUSSION

In the present study, tolerant individuals biased their reasoning about ethnic groups, while Thistlethwaite (5) using a similar method in 1948 reported intolerant bias alone. The present experiment conducted in 1955 in a social climate somewhat altered during the intervening years, used a completely Northern sample lacking highly prejudiced Ss. Thistlethwaite used Southern samples which presumably contained many highly prejudiced Ss. He compared the direction of reasoning bias in these with Northern samples which may have contained both intolerant and tolerant respondents. As a consequence, the tolerant bias he predicted for his Northern samples may have been attenuated. However, as Thistlethwaite selected his comparison groups on the basis of region rather than on standard attitude scales, more exact comparisons of these related studies cannot be made. Nevertheless, the findings of both studies taken together support the general assumption that individuals with opposing social attitudes distort their reasoning in directions congruent with these attitudes. Moreover, the findings from the present study challenge the conventional emphasis on veridicality in tolerant individuals. Apparently, tolerance is not always associated with superior reality oriented reasoning.

## SUMMARY

This study was undertaken to investigate further the relationships between ethnic attitudes

and objective reasoning about ethnic groups. It was argued that tolerant individuals would bias their reasoning in favor while intolerant individuals would bias their reasoning against ethnic groups. The findings revealed that tolerant individuals biased their reasoning significantly *in favor* of ethnic groups. However, intolerant individuals did not bias their reasoning significantly *against* ethnic groups. The lack of bias against ethnic groups was presumptively related to the absence of strongly prejudiced respondents in the sample. Biased reasoning on the part of unprejudiced individuals challenged the traditional assumption that tolerance is invariably accompanied by superior reality directed thinking.

## REFERENCES

1. ADORNO, R. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ALLPORT, G. W. *The nature of prejudice*. Cambridge, Mass.: Addison-Wesley, 1954.
3. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. I. *Theory and method*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 289-334.
4. SELL, D. E. *The evaluation aptitude test*. Chicago: Psychometric Affiliates, 1952.
5. THISTLETHWAITE, D. Attitude and structure as factors in the distortion of reasoning. *J. abnorm. soc. Psychol.*, 1950, 45, 442-458.

Received October 26, 1956.



## ERRATA

In the article "Perceptual Organization and Intelligence: A Further Study," by Clinton De Soto and H. Leibowitz, which appeared in the November, 1956 issue of this journal (*J. abnorm. soc. Psychol.*, 1956, **53**, 334-337), some incorrectly calculated probabilities of Type II errors were given. Recalculated probabilities are given in Table 1. These are probabilities of Type II errors for 2-tail *t* tests of the difference between mean vocabulary scores of high and low perceivers in case of certain hypothetical true differences in various experimental conditions and groupings, as explained in the original article. The probabilities in Table 1 are substantially greater than those given earlier, but low enough that a complete absence of significant *t*'s would be surprising if a large true difference existed. All *t*'s obtained were nonsignificant, and the average difference was in fact in the wrong direction.

It may be noted that 1-tail *t* tests might have been used, since the direction of the difference was hypothesized. One-tail *t* tests would have had lower probabilities of Type II errors than 2-tail *t* tests for the hypothetical true differences listed, and would also have yielded uniformly nonsignificant results.

TABLE 1

Condition	Difference	Probability of Type II Error
0.01 sec.	3.76	.65
0.01 sec.	5.20	.42
0.10 sec.	3.99	.56
0.10 sec.	5.53	.30
1.00 sec.	3.39	.57
1.00 sec.	4.70	.32
Above combined	1.87	.63
Above combined	2.53	.40
30 sec.	1.87	.72
30 sec.	2.53	.53

In the article by S. J. Korchin and S. Levine, "Anxiety and Verbal Learning," this Journal, 1957, **54**, 234-240, it was erroneously stated that Dr. Korchin is now at the Research Division, Columbus State Institute of Psy-

chiatry, Ohio State University. This statement applies to Dr. Levine. Dr. Korchin remains with the Institute for Psychosomatic and Psychiatric Research and Training, Michael Reese Hospital.

# ARE THERE GAPS IN YOUR FILES OF APA JOURNALS

---

## THEN HEAR THIS . . .

The American Psychological Association announces a sale during the period October 1957 through March 1958. Of the following journals, all *available* issues in the volumes for the years preceding 1951 will be offered at a price of only 50¢ (foreign, 60¢) per issue:

*American Psychologist*  
*Journal of Abnormal & Social Psychology*  
*Journal of Applied Psychology*  
*Journal of Comparative & Physiological Psychology* (1947-1950 only)  
*Journal of Consulting Psychology*  
*Journal of Experimental Psychology*  
*Psychological Abstracts*  
*Psychological Bulletin*  
*Psychological Index* (a few complete volumes, some shopworn)  
*Psychological Monographs*  
*Psychological Review*

---

Not *all* issues in *all* volumes are available. But—ORDER NOW before more back issues go out of print. From our available stock we will complete as much of your order as possible at this reduced price and for this limited period.

Delivery: 6 to 8 weeks

No dealer or quantity discounts

After this sale, for the years preceding 1948, journals will be available only on microfilm and microcard.

### Order from:

American Psychological Association  
Department BB, 1333 Sixteenth Street, N.W.  
Washington 6, D. C.

S  
A  
L  
E  
  
£  
  
B  
A  
C  
K  
  
I  
S  
S  
U  
E  
S



# CONSCIOUSNESS AS A REGULATORY FIELD: A THEORY OF PSYCHOTHERAPY

REX MADISON COLLIER

*VA Hospital, Jefferson Barracks, Missouri*

**T**WO PREVIOUS papers have outlined and partially developed a regulatory theory of consciousness. The first (3) was a brief initial statement which applied only to nonstress conditions; the second (4) developed the theory in terms of stress-defense dynamics and made applications to basic problems in psychopathology. The present paper proposes to interpret selected aspects of the process of psychotherapy in terms of these concepts.

Any adequate review of the literature on psychotherapy (e.g., Ellis [8]) reveals the growing tendency to eclecticism. In any field, eclecticism implies the need for a set of basic concepts that bring system, relatedness, and even unity to the common factors empirically selected from a variety of sources. Among the few current attempts to integrate the varied phenomena of psychotherapy in terms of a systematic set of concepts, the most notable have originated from Hullian learning theory (7, 12, 14, 15, 16). The chief difference between this type of theory and the kind here being developed is probably one of emphasis. Hullian theory emphasizes peripheral factors for the sake of accessibility and objectivity. The present theory places more emphasis on central factors with utilization of the biological concept of regulation. Some reduction in objectivity is risked for the purpose of attempting an organized approach which gives a more appropriate place to the unsuccessfully rejected concept of consciousness.

Any theory of how psychotherapy functions must be based on concepts of integration or the broader concept of regulation. The problem of psychotherapy may be stated as that of discovering and managing the conditions which facilitate the re-establishment of a more adequate degree of self-regulation. It is assumed that no definition or approach to psychotherapy is complete or sufficiently basic without including the concept of self-regulation in some form. The assumptions (a) that consciousness has a functional role in behavior and (b) that the concepts of field and regulation are essential to this function not only give new

meaning to the neglected but inescapable concept of consciousness but also lay the basis for developing and defining an interrelated group of dynamic concepts that seem to be basic in behavior.

While some writers (e.g., John Dewey [6]) have denied that consciousness can be successfully defined, for the purposes of this theory the following definition is proposed: Consciousness is an awareness field of biopsychological origin in which processes which may eventuate in action have increased chances of interaction and mutual modification. Thus, through such an awareness field, however simple and rudimentary or complex and comprehensive, the organism is provided with the means for a more flexible adaptation. In this sense, therefore, consciousness assumes a role which gives it great survival value in the evolutionary process. Since the problem of psychotherapy is to a large extent one of reorganization and readjustment, it is no mere coincidence that progress in reorganization is closely related to the degree to which the individual can allow himself to become aware of significant issues.

The regulatory effectiveness of the conscious field is dependent upon its magnitude which varies with the degree to which processes relevant to the current issues may be accessible to the field. A reduction of this kind of magnitude through blocking, resistance, or repression results in some degree of coincident reduction in the flexibility and regulatory adequacy of the conscious field. It is assumed that the more permeable the conscious field boundaries are to processes relevant to a current issue, the broader and more inclusive is the regulatory frame of reference.

The conditions producing variations in the regulatory effectiveness of the conscious field have been described in terms of five dynamically interrelated conceptualized variables. These are as follows: (a) stress, (b) defense system with primary and secondary components, (c) stress-tolerance thresholds, (d) ego strength, and (e) boundary permeability of the conscious field. These concepts, defined and

utilized in detail in a previous paper (4) lead to a central concept of stress-defense dynamics. In brief schematic form, the problem of psychotherapy in terms of the regulatory theory of consciousness and these related variables may be stated as follows:

1. The psychopathology which the individual (patient<sup>1</sup> or client) seeks to reduce is a product of the individual's system of defense interacting with experienced stress.

2. As a result of the stress-defense interaction, the regulatory field of consciousness becomes restricted to a degree that cripples its regulatory effectiveness. This allows the rejected segments of experience and of the self to produce biased and poorly managed impulses and behavior which often dominate responses to the further discomfiture of both the individual and his associates.

3. Assuming that the pathological processes are reversible to some degree, the first or primary essential in psychotherapy is the provision of a relationship designed to reduce the individual's need for his defensive system and coincidentally to increase his threshold of tolerance for stress. Therefore, some degree of dependence of the individual on the therapist for some period of time is necessary in most instances.

4. Communication between the individual and the therapist facilitates intrapersonal communication, which in turn increases chances for greater permeability of the restricting defensive boundaries to the conscious field.

5. Increases in ego strength—the ability to accept into consciousness potentially stressful material and to deal with it constructively—occur as the individual experiences acceptance by another who is significant to him. Thus, the individual learns to accept himself and thereby lays the basis for a more adequate degree of independence and self-regulation.

6. The rate at which rejected segments of experience and self can be returned to the regulatory conscious field depends upon all of these interrelated factors taken together, but

<sup>1</sup>The term "patient" is used here with special reservations. The traditional connotation that the patient is the passive recipient of things done to him and of orders carried out for him must be rejected. For psychotherapy to be effective, the patient must be an active participant in his own treatment. It is with this connotation that the term is used in this discussion. For discussions of this and related issues reference is made to Lindner (11) and Shakow (13).

initially the most important factor is the therapeutic relationship.

7. Increases in ego strength are of great importance, for they are associated with the following changes: (a) higher stress-tolerance thresholds, (b) reduction of need for the defense system, (c) increased permeability of the boundaries to the conscious field, and (d) increased self understanding or insight. Taken together, these results tend toward greater emotional and intellectual freedoms to move in directions more representative of the total personality.

It is assumed that there is no final or essential conflict between this theoretical approach and the established facts of learning. Some learning theories will emphasize particulars and their constancies, e.g., S-R, cues, verbal symbols, etc., while here the emphasis is on central regulation, field interaction, and the mutual modifiability of the interacting processes. Some may worry about the subjectivity aspect of the concept of consciousness, but it is important to note that even some behavioristic theories have much need for the intervening variable. The conscious field may be thought of, if one wishes, as a field of interacting intervening variables with regulatory significance for both the S and the R. This will emphasize, however, the "O" of Woodworth's S-O-R formula (17) and reduce the emphasis on S-R aspects. Often the essential difference between theories is the place of emphasis as one plays the scientific game of attempting to make a subject matter more intelligible.

During any attempt to escape the current eclecticism in psychotherapy and to arrive at a more integrative approach, a set of loosely related, often cumbersome, phenomena will be encountered. A selected group of these to which the regulatory theory of consciousness may be applied is as follows: communication and psychotherapy, free association, insight, effects of "uncovering" and interpretation, abreaction, and the relationship aspect of psychotherapy, and group relationship factors. While this is not an exhaustive list, it serves to illustrate how an ordering of the concepts may occur in terms of the regulatory theory of consciousness.

#### COMMUNICATION AND THERAPY

If the function of the conscious field is regulatory along with maintenance of flexibility of adaptation to immediate issues, then



the adequacy of this regulation is dependent upon what is available to the field. Cultural predispositions require, for the sake of clarity, a repetition of the point that this concept of consciousness is far broader than a synonym for reasoning, intellectualizing, or problem solving as these terms have been traditionally used. Western culture predisposes the individual to assume that it is "bad taste" to reveal strong feelings. The child who experiences strong emotion arouses anxiety in his elders and soon learns to fear and reject his own feelings. "Control" is thus acquired through an habituated exclusion of affect. The concept of consciousness that tends to be current is in part characterized by "flattened affect." Thus, the group predisposes itself to a culturally based incipient schizophrenia. To whatever extent the individual has denied himself the opportunity or privilege of experiencing his feelings, anxieties, dreads, elations, etc., to that extent he has become a more rigidly organized individual with fewer chances for these feelings to become flexibly related to other areas of his personality. A kind of intrapersonal communication has become interrupted. The function of communication in the process of psychotherapy is to make available to the conscious field, processes relevant to a given issue. Maximal adequacy of self-regulation is associated with maximal intrapersonal communication. Unacceptable experiences and feelings often become rejected and alienated from the main stream of awareness that tends to represent the self. From these rejected segments of experience, feeling, and self-evaluation, there are continued effects in the forms of anxiety, impulsive and compulsive desires, drives, irritations, and feelings of frustration that are poorly integrated with the main trends of personality. These rejected segments of feeling and experience are as rigid and unchanging as is their degree of dissociation from the conscious field. Re-association through improved intrapersonal communication is not only a goal of psychotherapy but is also recognized here as a goal of growth and increased maturity and hence should have an even wider application. One implied application is that the individual should learn to accept and respect the affect aspect of experience even though this may, at times, cause him considerable anxiety.

Since interpersonal communication tends to facilitate intrapersonal communication, psy-

chotherapy may be represented as a specialized and controlled means of facilitating intrapersonal communication. Barriers between dissociated segments of the personality tend to become more permeable as one continues to talk about the issues relevant to the dissociating processes. A simplified formula may be proposed as follows: *What one can talk about indicates what he can begin to accept, and what one can accept indicates what he can begin to manage.* Thus, a step is taken in the direction of a more adequate degree of self-regulation.

### FREE ASSOCIATION

The fundamental rule in traditional psychoanalytic procedures is that the analysand should say or represent verbally all that comes to mind, regardless of its nature. It can be seen immediately that, in terms of the regulatory theory of consciousness, the analyst attempts to utilize a basic device for facilitating the re-introduction of segments of experience to the conscious field. Verbalization heightens awareness, and the process of talking with reduced restriction encourages exploration and discovery of significant neglected and rejected aspects of experience. There are increased chances that a rejected aspect of experience reintroduced to consciousness will enter into mutually modifying relations with other processes currently available to the field. This can provide the discovered or uncovered aspect of experience with new context and a revised meaning. Hence, some increased degree of flexibility has been achieved. However, the invitation to a patient to use free association can be extremely threatening. It is an invitation in many instances to disturb, even to disrupt, a kind of equilibrium by which the individual has become accustomed to live. Insistence on the observance of the "fundamental rule" often places the analyst and the analysand in conflict, since the defense system of the analysand is invariably activated. Traditionally, the defense mechanisms were viewed as more or less nuisance devices that prevented the analysis from progressing smoothly toward an early successful conclusion. The approach of "Off with his defenses" often has subtle similarities to the Red Queen's command of "Off with their heads!"

No therapy is adequate that does not respect the individual's defense system as a basic aspect of his personality. Free association, be-

cause of its potencies, is both a useful and a hazardous technique. To invite the individual to free associate is to invite him to lay aside his defenses and accept back into awareness that which he has found habitually unacceptable and stressful. In terms of stress-defense dynamics, free association may be either harmful or beneficial. It is hypothesized, in terms of the present theory, that free association is therapeutically successful under the following circumstances: (a) if the individual's defenses operate well enough so that stressful material is fed back to the conscious field at a rate that does not exceed ego strength as here defined; (b) if the relationship with the therapist is sufficiently supportive so that there is some increase in ego strength; and (c) if the therapist's interpretations do not release anxiety that exceeds the patient's stress-tolerance threshold. Free association can be less than successful, or even harmful, if the patient's secondary defenses are breached before there has occurred sufficient development of ego strength so that secondary defenses may be less necessary to the patient. If the stress of free association in conjunction with too penetrating interpretations robs the individual too quickly of his secondary defense components, he will revert to primary defense. The ensuing results depend upon the seriousness of the problem. The patient may in some of the less serious instances simply leave therapy with less accessibility to future attempts, or, if he stays in therapy, he may show psychotic trends. In the more serious instances, a psychotic episode may occur. It is hypothesized that, were adequate records of psychotic breaks during analyses available, it would be found (a) that the patient initially had an inadequate secondary defense elaboration and depended chiefly on primary defenses—a dynamic arrangement which was not clearly recognized by the therapist—or (b) that what secondary defenses he had were interpreted away before ego strength had improved sufficiently to tolerate the added stress.

The couch and a reclining position are usually thought of as related to the technique of free association. Placing the patient on a couch where he cannot face his therapist can be damaging if the initial ego strength is not high enough to accept the stress involved in such a nonsupportive relationship. For a pa-

tient with weak ego strength, which tends to go with predispositions to utilize primary defense, to be on a couch with an aloof analyst out of sight and to be requested to free associate about his problems is to take a long step in the direction of precipitating a psychotic episode.

Free association is a powerful technique for facilitating the reacquisition of experiences, feelings, and meanings to the conscious regulatory field, but the speed of reacquisition must be compatible with the balance between stress-tolerance thresholds, or thresholds of defense reaction, and current ego strength. The process traditionally referred to in psychoanalysis as "making the unconscious conscious" needs extensive review in terms of the stress-defense dynamics which are here being outlined.

### INSIGHT

From what has thus far been written, it should be clear that the concept of insight can be useful, but it must be far broader in psychotherapy than the insight of nonemotional, intellectual, problem-solving behavior. Insight in psychotherapy is based on the acceptance into consciousness of hitherto neglected or rejected aspects of experience under conditions of sufficient emotional permissiveness for these aspects, as well as other relevant aspects of experience, to interact so that new and more comprehensive meanings emerge. Insights are, therefore, limited (a) if the boundaries to the conscious field are relatively impermeable to thoughts, feelings, and tendencies relevant to the problem, (b) if ego strength is low, and (c) if experienced stress is currently high enough to keep the primary component of the defense system active. The sudden, dramatic insights, described in the textbooks with misleading frequency, are simply one of the rarer aspects of the broader principle of intrapersonal communication referred to earlier. To be "open-minded" (the concept of the permeable barrier to the conscious field) toward one's experiences and feelings and to what these have meant to one, is a basic objective not only in psychotherapy but in the maintenance of mental health for everyone. If this condition of the conscious regulatory field is maintained, then insights may occur fairly frequently, but few of them are likely to be dramatic with regard to one's emotional life, since his adjust-



ment problems are kept up-to-date and there are no significant areas of experience that remain rejected or alienated.

While the facilitation of insightful reactions is an important goal in psychotherapy, the more fundamental objective is to provide the individual an opportunity gradually to accept and to deal realistically with aspects of himself and his experience that heretofore he has had to neglect or reject. The reaction of insight can be thought of as a criterion that such intrapersonal communication is occurring. However, insights must be more than logical; they must also represent an acceptance of the emotional meaning of experience. The logical insight by itself may indicate that the more affective aspects of experience are still being rejected. Logical or intellectual insight tends to modify behavior chiefly under nonstress conditions. Where the problem involves threat and anxiety sufficient to activate defenses, the insight must be affectively toned and comprehensive enough to indicate a real acceptance of the basic threat issue. The terms and processes of logical problem-solving are often less than adequate for behavior under stress.

If the fundamental goal of psychotherapy is to re-establish a more adequate degree of self regulation, then the techniques used must facilitate through interpersonal communication an improved intrapersonal communication. The broader concept of insight as used here becomes one criterion for movement toward this goal.

#### THE UNCOVERING OR INTERPRETIVE CONTINUUM

When a therapist interprets or invites a patient to explore his experiences, the procedure is sometimes referred to as an "uncovering" technique. How a therapist's response may be scaled according to the degree of uncovering represented has been outlined in previous studies (2). The concept of scaling the depth of interpretive responses of the therapist is of great importance in reference to the stress-defense dynamics. From this a clearer rationale seems to appear for determining the degree of depth of interpretive response of the therapist appropriate at any particular time for the patient. The question may be asked whether the therapist should restrict himself to listening, to repeating significant words or phrases, to clar-

ifying or reflecting, or whether he should include greater depth in some of his responses. If the objective of therapy is aiding the patient in achieving an improvement in self-regulatory ability, then it must be assumed that he needs to accept into the conscious field and to integrate constructively into action patterns some aspects of himself and of reality around him that heretofore he has rejected or neglected. The way in which he presents his problem at any given time reveals his capacity (ego strength) to deal with the problem. What he leaves out or attempts to avoid reveals, in part, his use of primary defense (not simply resistance). How he rationalizes, compensates, and utilizes various symptom patterns reveals not only his need for secondary defense but also his capacity to blunt the sharpness of stress factors so that he can still deal with reality to some extent (4). It often occurs in the beginning of a therapy relationship that the patient is quite restricted in his approach to the problem and in what seems to him as relevant. The therapist's technique in managing the relationship is a first step in reducing stress which often aids in some coincident increase in ego strength. Material relevant to the problem, but held at a superficial level of rejection, may still not be verbalized. Stress effects are often obvious, but, if stress-tolerance thresholds are not greatly exceeded, the patient may, even so, deal with the material integratively. Any interpretation that brings in material that the patient can neither accept nor deal with constructively, that reactivates primary defenses, or that stimulates symptom formation may be said in general to exceed the therapeutically appropriate. For the therapist to insist on an interpretation which involves the risk that the patient may feel less accepted or less understood is to retard the therapeutic processes as defined here. The therapist may be as penetratingly interpretive as the current stress-defense dynamics predispose the individual to accept and to utilize such interpretations. This contention involves the assumption that the interpretation has some degree of validity that may be sifted out by continued exploration.

A persistent hazard of the more penetrating interpretation is that it can rob the patient of the ego strength-building experience of self-discovery. If the purpose of therapy is to improve the quality and adequacy of self-

regulation, then there should be maximum opportunity for the patient to make his own discoveries and his own applications. The appropriate technique for the therapist often grows out of the question put to himself, "How can I set the situation so that the patient can recognize a feeling, relationship, meaning, etc., and retain the feeling that he discovered it for himself?" While this cannot always be done, it should occur sufficiently often, or the therapist is not providing an opportunity for the self-regulatory capacities of the individual to grow. This approach not only avoids overdependence on the part of the patient, but it also represents an area in which psychotherapy and teaching should have greater overlap.

To make a sharp distinction between supportive (often called "covering") and exploratory (or "uncovering") therapy has become relatively frequent. In terms of the stress-defense dynamics which have been discussed, this kind of dichotomy is undesirable. To talk about one's problems is to uncover threatening aspects in some degree. How much revelation is appropriate depends on the relationship between therapist and patient, the degree of ego strength available to the patient, and the balance between primary and secondary segments of the defense system. The supportive-uncovering dichotomy should be thought of as a depth continuum. Currently, the therapist must depend on his clinical judgment as to how exploratory he should be or how exploratory he should allow his patient to be or at what level on the depth continuum it is appropriate to work. Depth can be defined here in terms of the typical degrees of hesitation and resistance encountered in the social group in discussing issues of dynamic importance to them. Such a scale has yet to be devised, but it is experimentally feasible and could have a variety of uses. In any case, some degree of depth is represented by any discussion. Therefore, it is important to point out that no therapist can be supportive in any adequate or relevant manner until his patient has provided some exploration of his problems. To attempt to be only "supportive" in the traditional sense is often to leave the patient with a feeling that he is not well understood or even that he has been somewhat rejected.

## ABREACTION

Abreacting a traumatic or stressful experience has been recognized as having both therapeutic and nontherapeutic effects. Grinker and Spiegel (10) have said that abreactions under pentothal are not necessarily therapeutic, but may provide a beginning of useful insight. They have also said that abreactions in a bar under the influence of alcohol are nontherapeutic. Dollard and Miller (7) propose that the difference between therapeutic and nontherapeutic abreactions may be understood in terms of the concepts of a conditioned response theory of learning, e.g., reward and punishment, reinforcement and extinction, and generalization. How would stress-defense dynamics conceptualize this issue?

Abreaction, as a reliving of a past experience, provides an opportunity, under conditions of increased permeability of conscious boundaries, for a rejected segment of experience to be re-introduced to the conscious field and to be assimilated. By assimilated is meant (a) that the experience has an increased degree of acceptance and (b) that it is provided an opportunity to enter into mutually modifying relations with other processes also available to the conscious field. That is to say, in the new context it now has changed meaning. Accepting the segment of experience which has been persistently alienated implies a reduction in defense operations at the time. This can occur in spite of the stressful nature that abreaction seems at times to have only if the patient has had an increase in ego strength. Such a result grows out of the degree of confidence he has in the therapist and the treatment procedures. It is also therapeutically possible for an abreaction to be beneficent apart from a therapist or institutions, provided the individual can accept the products and deal with them with at least some degree of reduced defensiveness. This can occur only when some current set of factors has provided an increase in ego strength as defined here.

Abreaction can be nontherapeutic or even detrimental if the individual cannot accept the revived segment of experience. Revival may be brought about by means of drugs, free association, hypnosis, or other methods. Since the stress generated will activate the defense system which has maintained the rejection or



alienation, the personality may reconstitute itself along habitual lines. This may be thought of as simply a nontherapeutic result. The experience may be detrimental if the situation in which abreaction occurs is now also rejected. The therapist, the institution, or the treatment procedure may be rejected and the patient is then less accessible than he was previously. The area of trauma has become enlarged. In general, to be therapeutically successful, abreactions should occur only when there is a good relationship of confidence and mutual acceptance between therapist and patient. How soon this occurs in therapy depends on the personalities of the people involved; it may occur in the first interview or it may never occur.

#### THE RELATIONSHIP ASPECT OF PSYCHOTHERAPY

The importance of establishing and managing a relationship between therapist and patient that may be utilized for therapeutic ends has long been recognized. This relationship, the subject of many treatises, is often described in language vague and mystical. It is hoped that a step in the direction of a more operational approach is accomplished by invoking stress-defense dynamics in this issue. When the therapist is accepting and relatively nonjudgmental toward his patient in managing this relationship, he is (a) reducing a sense of stress and (b) reducing to some degree the patient's need for defense. With less need for defense, there is some increase in the permeability of the barriers to the conscious field. Coincidentally, the patient perceives not only the therapist's acceptance but his imperturbability or lack of surprise at the "terrible" things the patient tells him. This apparent strength of the therapist is a support to the patient's own ego strength. Thus, on this basis also, there is some increase in the patient's ability to accept into consciousness potentially stressful material with increased chances of dealing with it constructively. Not only has a better basis for interpersonal communication been laid, but some facilitation in intrapersonal communication may occur. Whole therapies are based on providing an individual with an opportunity to talk out his problems in a non-threatening relationship. Some therapists have been so convinced of the effectiveness of the

relationship itself that they have tended to renounce analytically derived free association and interpretation. One may hypothesize in terms of stress-defense dynamics that the aspects of relationship described above are basic to any additional elaboration of technique. According to Fiedler (8), the "experts" of various schools tend to have greatest similarity at the point of managing the relationship aspect of therapy.

The facets of the relationship between therapist and patient often become both subtle and complex. In good relationship therapy, the patient utilizes the relationship for a broader and more complete expression of his own rejected feelings and tendencies. The therapist utilizes the relationship to aid the patient in becoming increasingly aware of the nature of his problems. Thus, according to the regulatory theory of consciousness, to become aware under conditions of reduced threat is to increase chances of improved self-regulation.

#### GROUP RELATIONSHIP FACTORS

The concept of ego strength has been defined as the ability to accept into awareness potentially stressful material and to deal with it constructively. It is hypothesized further that ego strength, as here defined, is largely a product of a sense of security and a sense of belonging in a significant way to a group of other human beings. If this part of the theoretical frame of reference is accepted, several inferences are immediately derivable. Certainly one of the first steps in individual therapy is the extension of real acceptance by the therapist to the patient. Some feeling of being accepted and at least somewhat secure in the two-group relationship is necessary on the part of the patient or therapy cannot even get started.<sup>2</sup> An increase in group participation on the part of the patient during therapy is both diagnostic and therapeutic. This is an indication of improvement and a utilization of relationships that increase chances for further improvement.

If ego strength is a product of a sense of security and belonging, then a look at home and community relationships—our typical cultural pattern—is not very reassuring. One needs only to think of the multitude of ways

<sup>2</sup> Cf. Bobbitt and Clausen (1).

by which the typical community rejects, neglects, avoids, or makes invidious distinctions among its members to realize how ego-crippling experiences can hardly be avoided. The positive side of this community picture often looks good, but most communities unwittingly set situations that result in tragic damage to some of its members. This is of extreme importance since the increase of population, together with a loss of frontiers, forces people to live in closer and closer contacts. People must learn to live together constructively or they will increasingly destroy each other.

#### SUMMARY

Psychotherapy is defined as a set of techniques, all of which should have as their common purpose the progressive re-establishment in the individual of a more adequate degree of self-regulation. The pathologies of self-regulation have been ordered to a set of concepts referred to as stress-defense dynamics. The essential problem in psychotherapy is the recognition and appropriate manipulation of these variables in order to effect a maximum degree of reversal of pathological processes. This systematizing approach has been applied to a selected cluster of phenomena related to psychotherapy. These are as follows: communication and psychotherapy, free association, insight, effects of "uncovering" and interpretation, abreaction, the relationship aspect of psychotherapy, and group relationship factors. It is assumed that the current style of being eclectic in psychotherapy demands that certain basic concepts be systematically identified so that the recognized common factors in psychotherapy shall have an organized meaning. The regulatory theory of consciousness and the derived concepts of stress-defense dynamics are offered as a step in this direction.

#### REFERENCES

1. BOBBITT, J. M., & CLAUSEN, J. A. Psychotherapy and its public health implications. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald Press, 1953. Pp. 171-201.
2. COLLIER, R. M. A scale for rating the responses of the psychotherapist. *J. consult. Psychol.*, 1953, **17**, 321-326.
3. COLLIER, R. M. Outline of a theory of consciousness as a regulatory field: Preliminary statement. *J. Psychol.*, 1955, **40**, 269-274.
4. COLLIER, R. M. Consciousness as a regulatory field: A theory of psychopathology. *Psychol. Rev.*, 1956, **63**, 360-369.
5. COLLIER, R. M. A definition of treatment. *The Amer. J. Psychother.*, 1957, **11**, 45-55.
6. DEWEY, JOHN. *Psychology*. New York: Amer. Book Co., 1891, p. 2.
7. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950, p. 372.
8. ELLIS, Albert. New approaches to psychotherapy techniques. *J. clin. Psychol.*, Suppl. Monogr. 1955, **11**, 208-260.
9. FIEDLER, F. E. A comparison of therapeutic relationships in psychoanalytic, non-directive, and Adlerian therapy. *J. consult. Psychol.*, 1950, **14**, 436-445.
10. GRINKER, R. R., & SPIEGEL, J. P. *Men under stress*. New York: Blakiston, 1945, p. 392.
11. LINDNER, R. M. Who shall practice psychotherapy? In M. H. Krout (Ed.), *Psychology, psychiatry, and the public interest*. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 148-160.
12. MOWRER, O. H. Learning theory and the neurotic paradox. *Amer. J. Orthopsychiat.*, 1948, **18**, 571-610.
13. SHAKOW, D. Psychology and psychiatry: A dialogue. *Amer. J. Orthopsychiat.*, 1949, **19**, 191-208; 381-396.
14. SHAW, F. J. A stimulus-response analysis of repression and insight in psychotherapy. *Psychol. Rev.*, 1946, **53**, 36-42.
15. SHOEN, E. J. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, **46**, 366-392.
16. SHOEN, E. J. A learning-theory interpretation of psychotherapy. *Harvard educ. Rev.*, 1948, **18**, 129-145.
17. WOODWORTH, R. S. *Psychology*. New York: Henry Holt, 1929, p. 226.

Received February 1, 1957.



# THE EFFECTS OF FEAR AROUSAL AND SUPPRESSION OF FEAR UPON SOCIAL PERCEPTION

SEYMOUR FESHBACH AND ROBERT D. SINGER

*University of Pennsylvania*

THE PRESENT paper is concerned with some factors influencing the presence and extent of "projection" in social perception. The term "projection," of psychoanalytic origin, has accumulated diverse connotations (4, 5) and has acquired considerable explanatory power at the cost of precise analysis of the phenomena which it presumably explains. Thus, the paranoid who is defending against persecutory agents, the miser who perceives his neighbors as equally avaricious, and the fearful child whose world is full of danger and threat are all said to be "projecting."

The present experiment is delimited to the study of one of the psychological processes underlying the various forms of projection; namely, the influence of affect upon perception. Our basic proposition is as follows: when a particular affective state is aroused, there is a tendency for the affect to become connected to contemporaneous percepts and ideation. One might describe this process as the infusion of cognition with affect, a process very similar to Heinz Werner's concept of "physiognomic" perception (6). Freud (3) has also described a process of this kind and Bellak (1), in his theoretical analysis of apperceptive distortion, refers to the role of the mood of the perceiver.

Infusion is conceived of as a basic, primitive organismic process that is especially characteristic of infants and young children. It is assumed that the infusion tendency persists to some degree in adults although greatly modified by learning. One has to learn not to project, as it were; that is, learn to discriminate between one's own feelings and those of other objects in the environment. Repression that tends to reduce the effects of higher-order learning (2) should result in greater projection or infusion. If a subject is unaware of or denies his feelings, then he is unable to employ an important cue to which he has previously learned to make discriminating responses.

These considerations lead to the following specific hypotheses:

1. Under conditions of fear arousal, a subject perceives (judges) a stimulus person as sig-

nificantly more fearful and anxious than under neutral affective conditions.

2. Under conditions of fear arousal, subjects encouraged to suppress recognition of their emotional reactions perceive a stimulus person as significantly more fearful and anxious than subjects encouraged to acknowledge their emotional reactions.

## METHOD

### *Subjects*

Sixty male volunteers from introductory psychology classes at a large metropolitan university participated as subjects. The Ss were randomly assigned to one of the following experimental treatment groups: Control (C), Fear-Expression (FE) and Fear-Suppression (FS).

### *Administration of Film and Fear Stimulus*

The Ss were seen individually. Each S was told that the experiment dealt with the effects of distraction upon the accuracy of one's judgment of other people. They were informed that a film of a young man's performance on a number of different tasks would be shown to them, and that after presentation of the film, they would be required to judge the personality of this man. Following these instructions, electrodes were attached to the left ankle of Ss in the Fear groups. Electric shocks, administered in the guise of distracting stimuli, were employed to induce fear. The strength of electric shock administered to each S was determined by gradually varying the intensity of the shock until S reported that the shock was painful. The film was then presented, and eight shocks were administered at varying intervals while the Ss witnessed the film. At the same time, they were required to remove a set of pins from the O'Connor Finger Dexterity Board, a task that could be readily accomplished with only occasional glances at the board. A similar procedure was used for the Control group with the important exception that shock was omitted.<sup>1</sup>

### *Manipulation of Fear Expression and Fear Suppression*

The instructions intended to encourage expression of fear or suppression and denial of fear were given

<sup>1</sup> In order to determine whether the pin removal task had any effect upon the Ss' reactions to the film, half the Ss in the Control group were not given this task to perform. The means and variances of the subsequent judgments of the two Control subgroups were practically identical, and consequently the subgroups were combined in analyzing the results.

prior to the showing of the film. The Ss in the Fear-Expression (FE) group were told: "Many people are disturbed by the shock. In order to perform most efficiently, the best thing to do is to be aware of and admit your feelings. If you freely express your feelings, your judgments will be more accurate." The Ss in the Fear-Suppression (FS) group were given contrasting instructions as follows: "Many people are disturbed by the shock. In order to perform most efficiently, the best thing to do is keep your mind off your emotional reactions and not think about them. Try to forget about your feelings and concentrate on the task; by so doing, your judgments will be more accurate."

### *Measurement of Experimental Effects*

Subsequent to witnessing the film, Ss were administered a questionnaire regarding the personality characteristics of the stimulus person.<sup>2</sup> Five alternative choices were offered for each item in the questionnaire. The following four indices reflecting the evaluation of different personality attributes of the stimulus person were derived from the questionnaire responses:

*Indirect Fear Judgment score.* This index, based on 24 items, reflects the degree to which the stimulus person was judged as fearful or anxious in situations other than that depicted in the film. Sample questions: "How much tension do you think he usually works under?"; "How often do you think his sleep is fitful and disturbed?"; "In comparison to most people, how much does physical pain upset him?"

*Direct Fear Judgment score.* This index, based on 4 items, indicates the degree to which the stimulus person was perceived as fearful and anxious in the actual test situation depicted in the film. In making direct fear judgments, the S has more stimulus information available than in making indirect fear judgments. Sample questions: "How worried do you think he was while working on the various tasks?"; "How frequently did he seem to be anxious?"

*Indirect Aggression Judgment score.* This index, based on 13 items, reflects the degree to which the stimulus person was perceived as a generally aggressive individual. Sample questions: "How frequently do you think he loses his temper?"; "When he gets angry, how often do you think he takes it out on somebody else?"

*Negative Personality Judgment score.* This index, based on 15 items, reflects the extent to which unfavorable personality characteristics other than anxiety or aggressiveness were attributed to the stimulus person. Sample questions: "How much reliance do you think one can place upon his word?"; "How tolerant is he likely to be of other people's mistakes?"

A further control was introduced during the completion of the questionnaire in that the shock apparatus was detached from half of the Ss in each of the Fear-

Arousal groups while for the remaining half, the shock apparatus was left intact and these Ss were given three shocks during this period. Since this condition did not result in any significant differences in subsequent judgments, these two subgroups are combined in the presentation of the results.

### *Apparatus*

The S was seated at a table situated nine feet from the screen on which the film was projected. The electric shocks were administered by means of a high impedance 60-cycle inductive generator. The film, which lasted approximately 7 minutes, depicted a young man's performance on a card-sorting task, the O'Connor Wiggly-Block Test, and the O'Connor Tweezer Dexterity Test.<sup>3</sup> Neither sound track nor film titles were employed. The examiner administering the tests was not shown in the film except for a brief glimpse of his hands when he changed the test materials. The stimulus person, who had had previous practice with the various tasks, was instructed to complete each task as quickly as he could.

### RESULTS

The means for all groups on each of the questionnaire subscales are presented in Table 1. The results of an analysis of variance of the Indirect Fear Judgments are given in Table 2. The *F* ratio is significant at less than the .01 level, the FS group having the highest Indirect Fear Judgment scores and the Control group, the lowest. The differences between each of the Fear-Arousal groups and the Control group are significant at the .01 level. The difference between the Fear-Suppression and the Fear-Expression groups, although in the predicted direction, is not statistically significant. However, examination of the individual item scores revealed a substantial difference between the two groups in the extent to which they used the most extreme fear category (alternative 5) in judging the stimulus person. As Table 3 indicates, 60 per cent of the FS group in contrast to 25 per cent of the FE group judged the stimulus person as extremely fearful on at least one of the items. This difference is significant at the .05 level of confidence. Since the differences between the Fear groups and the Control group may be taken as measures of projection, the results, in support of the first hypothesis, indicate that Ss experiencing shock project

<sup>2</sup> A copy of this questionnaire has been deposited with the American Documentation Institute, Order Document No. 5324 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C. remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

<sup>3</sup> The authors wish to express their gratitude to Dr. William Nicholson who served as the principal character in the film and to the staff of the F. W. Taylor Management Laboratory of the Wharton School, University of Pennsylvania, who gave their time and facilities freely in processing the film.



their fear onto the stimulus person. There is also some support for the second hypothesis; namely, that suppression of fear facilitates the projection process.

Included in the questionnaire were four items comprising the Direct Fear Judgment Index which required *S* to judge how worried or anxious the stimulus person appeared during the test situation depicted in the film. Analysis of these data yielded results in the same direction as the previous findings, but the differences were less striking. The data, based on only a few items, were much more variable than those for Indirect Fear Judgment and, as Table 4 indicates, the *F* ratio obtained is not significant. Comparisons between groups based on each of the four items comprising the Direct Fear Judgment Index revealed no significant differences between the FE and C groups and between the FE and FS groups. However, on one of these items, the difference between the FS and C groups is significant at the .05 level, on two of the items, at the .06 level, and, on the remaining item, the difference is not significant.

The *Ss*' perception of the aggressiveness of the stimulus person was considered in the study because of the possibility of so-called complementary projection and also because the same operations which aroused fear may also have aroused aggression. An analysis of variance of the Indirect Aggression Judgment scores is presented in Table 5. The *F* ratio is significant at the .01 level, the FS group having the highest aggression perception scores and

TABLE 2  
ANALYSIS OF VARIANCE OF THE INDIRECT FEAR JUDGMENTS

Source	Sum of Squares	df	Mean Square	F
Between groups	3546.55	2	1773.3	8.04**
Within groups	12573.65	57	220.6	

\*\* Significant at the .01 level.

TABLE 3  
USE OF EXTREME FEAR CATEGORIES ON INDIRECT FEAR JUDGMENT SUBSCALE

Group	Made 1 or More Extreme Fear Judgments	Made no Extreme Fear Judgments
Fear suppression (FS)	12	8
Fear expression (FE)	5	15

Note.— $\chi^2 = 5.8$ ;  $p < .05$ .

TABLE 4  
ANALYSIS OF VARIANCE OF THE DIRECT FEAR JUDGMENTS

Source	Sum of Squares	df	Mean Square	F
Between groups	70.23	2	35.1	2.09
Within groups	956.10	57	16.8	( $p < .20$ > .10)

TABLE 5  
ANALYSIS OF VARIANCE OF THE AGGRESSION JUDGMENTS

Source	Sum of Squares	df	Mean Square	F
Between groups	633.0	2	3.7	5.76**
Within groups	3144.0	57	55.2	

\*\* Significant at the .01 level.

the C group the lowest. Thus, it appears that aggression as well as fear was projected in this situation.

In view of this finding, one might conjecture that the judgments of the Fear-Arousal groups are largely a reflection of a generalized negative attitude toward the experimenter and the experimental situation. If this were the case, one would expect the shocked *Ss* to be more likely than the C group to attribute negative personality characteristics (other than aggression or fear) to the stimulus person. However, an analysis of variance of the Negative Per-

TABLE 1  
COMPARISON OF MEAN SUBSCALE SCORES

Subscale	Group			$\bar{x}$ FE vs. C	$\bar{x}$ FS vs. C	$\bar{x}$ FS vs. FE
	Control (C) ( <i>N</i> = 20)	Fear Expression (FE) ( <i>N</i> = 20)	Fear Suppression (FS) ( <i>N</i> = 20)			
Indirect fear (24 items)	55.9	68.9	74.2	2.7**	3.9**	1.1
Direct fear (4 items)	8.5	9.7	11.2	.9	2.1*	1.2
Aggression (13 items)	30.2	34.4	38.1	1.8	3.3**	1.6
Negative Personality* (15 items)	47.4	44.6	46.3	—	—	—

\* In the case of the Negative Personality Judgment items, the lower the mean, the more unfavorable the judgment.

\* Significant at the .05 level.

\*\* Significant at the .01 level.

TABLE 6  
ANALYSIS OF VARIANCE OF THE NEGATIVE PERSONALITY  
JUDGMENTS

Source	Sum of Squares	df	Mean Square	F
Between groups	44.6	2	22.3	.56
Within groups	2266.8	57	39.8	

sonality Judgment scores presented in Table 6 indicates that differences among the three groups are slight and insignificant. These data indicate then that differences in perception between the experimental groups and the control group are specific to fear and aggression and cannot be attributed to a generalized "halo" effect.

While not directly relevant to the hypotheses under study, the correlation between judgments of fearfulness (Indirect) and judgments of aggressiveness is of some interest. The Pearson  $r$  calculated for all groups combined is .68 which is significant at the .01 level. The individual correlations for the C, FE, and FS groups are .65, .43, and .72, respectively, and are also statistically significant. The fact that even in the control group there is a substantial relationship between the tendency to perceive an individual as anxious and the tendency to perceive him as aggressive indicates that the obtained correlations are not solely a result of the experimental operations. An interpretation of this correlation as a reflection of a favorable-unfavorable attitude toward the stimulus person is suggested by the significant negative correlation of  $-.51$  obtained in the Control group between the Indirect Fear Judgment scores and the Negative Personality Judgment scores (the lower the score on this scale, the less favorable the judgment).

#### DISCUSSION

The results, when considered as a whole, lend substantial support to the hypothesis that the arousal of fear results in a tendency to perceive another person as fearful and anxious, at least within the limits of the experimental operations used to arouse fear and to measure its consequences. In evaluating the significance of these experimental results, it is important to note that although all Ss were oriented toward making accurate personality judgments of the stimulus figure, relatively few relevant cues

were available for making such judgments. In the case where more relevant cues were present, such as in those instances where Ss were asked to report the actual behavior or feelings of the stimulus person in the film situation, the effects of fear arousal, and particularly suppression of fear, are less striking than where indirect anxiety and fear were being judged. It has already been suggested that this difference may be a function of the difference in number of items and, hence, in the reliability of the scales used. The available data do not permit an assessment of the relative roles of statistical reliability and the constraints imposed by the stimulus, but there is sufficient theoretical and empirical basis for expecting greater perceptual distortion, or in this context, greater infusion, with reduction in stimulus information.

Of considerable interest are the results bearing upon the hypothesis that suppression of fear facilitates the tendency to project fear onto other social objects. On the measures of both Indirect Fear Judgment and Direct Fear Judgment, the Fear-Suppression group attributes a greater degree of fear and anxiety to the stimulus person than does the Fear-Expression group. When the two groups are compared with respect to the use of extreme fear categories, the difference between them is statistically reliable. These data, as well as those for Indirect Aggression Judgment, are then consistent with the hypothesis that suppression of affect facilitates projection. However, the effects of this variable are not as clear-cut as the effects of fear arousal, *per se*, and the results do not unequivocally support this hypothesis.

It must also be noted that the hypothesized explanation of the effects of suppression in terms of avoidance of cues essential for discrimination between the affective experience of the perceiver and the affect of the perceived object, is only one of several that can be offered to account for the experimental findings. Thus, although both experimental groups were reassured that many subjects are disturbed and upset by the experimental conditions, the Suppression group may have felt additional anxiety because of inability to suppress their feelings and thereby conform with the instructions. Also, greater conflict in this group may have contributed to a greater level of tension. Thus, the differences in amount of



projection between the Fear-Suppression and Fear-Expression groups may have been functions of differences in the amount of fear and aggression experienced. Data pertinent to the

similar experimental conditions, pictures of children were used as stimuli, supplementary projection would be obtained; that is, the perception of the affective state of the stimulus affective

size that  
a social  
interacting  
state of

igned to  
pression  
hypothesis  
a fear-  
project  
cts, and  
real reac-  
jection.  
ne of the  
ps: Fear  
Control.  
intervals  
witnessed  
f a num-  
tation of  
up were  
ey might  
contrast,  
were en-  
air emo-  
witness-  
shock.  
Ss com-  
the per-  
al in the

:  
e Fear-  
ilus per-  
nd also  
ial than

suppres-  
s were  
n. The  
ir mean  
ally sig-  
greater  
roup in  
up used

TABLE 6  
ANALYSIS OF VARIANCE OF THE NEGATIVE PERSONALITY  
JUDGMENTS

Source	Sum of	df	Mean	F
Between				
Within g				

were available for making such judgments. In the case where more relevant cues were present, such as in those instances where Ss were asked to report the actual behavior or feelings of the

sonality  
indicate  
groups :  
indicate  
between  
control  
and ca  
"halo" :

While  
under :  
ments c  
of aggr  
Pearson  
.68 whi  
individu  
groups :  
are also  
even in  
relations  
an indiv  
perceive  
obtained  
the expe  
of this c  
unfavor:  
person i  
correlati

ive

The re  
substant  
arousal  
another  
within t  
used to  
quences  
experim  
that alt  
making  
stimulus



projection between the Fear-Suppression and Fear-Expression groups may have been functions of differences in the amount of fear and aggression experienced. Data pertinent to the issue of the mechanism responsible for the effects of suppression could be furnished by experimental research involving positive emotions such as sympathy and joy. However, whatever the mechanism by which suppression and other variables may influence projection, it is important that one first establish the parameters of the projection or infusion process. Thus, the present results indicate that while suppression or "repression" of affect is not essential in order for projection to take place, it enhances the degree of projection present.<sup>4</sup>

It is instructive to examine the differences between the results of this study and those of Murray's classic study of projection (4). Although both studies deal with the effects of fear, we obtained supplementary projection and possible complementary projection while Murray obtained complementary projection only; that is, the children of Murray's study perceived the stimuli as malicious rather than fearful. An obvious explanation of this difference lies in Murray's procedure which involved rating of the pictures only with regard to maliciousness. However, we suggest that even if Murray had obtained judgments of fearfulness, changes in the ratings of fearfulness would have been less marked than changes in the ratings of maliciousness. A critical variable in our view is the relationship between the perceiver and the stimulus person. In Murray's study, the children were asked to judge pictures of adult males. Adults similar to the latter are likely to have had threatening roles in previous situations in which children have experienced fear, and consequently their faces were perceived as malicious. In terms of our theoretical framework, we would predict that if, under

similar experimental conditions, pictures of children were used as stimuli, supplementary projection would be obtained; that is, the perception of the affective state of the stimulus person would directly match the affective state of the perceiver.

These considerations serve to emphasize that the final judgment or perception of a social object is the outcome of a number of interacting variables, one of which is the affective state of the perceiver.

#### SUMMARY

The present experiment was designed to study the effects of fear arousal and suppression of fear upon social perception. It was hypothesized that individuals subjected to a fear-producing situation would tend to project their feelings upon some social objects, and further, instructions to inhibit emotional reactions would increase the amount of projection.

The Ss were randomly assigned to one of the following experimental treatment groups: Fear Expression, Fear Suppression, and Control. Shock was administered at varying intervals to Ss in both Fear groups while they witnessed a film of a young man's performance of a number of different tasks. Prior to presentation of the film, Ss in the Fear Expression group were encouraged to recognize any feelings they might experience because of the shock. In contrast, Ss in the Fear Suppression group were encouraged to suppress and inhibit their emotional reactions. The Control group also witnessed the film but did not receive any shock. Subsequent to witnessing the film, all Ss completed a questionnaire dealing with the personality characteristics of the individual in the film.

The following results were obtained:

1. The Fear-Suppression and the Fear-Expression groups perceived the stimulus person as a significantly more fearful and also significantly more aggressive individual than did the Control group.

2. The differences between the Fear-Suppression and the Fear-Expression groups were consistently in the predicted direction. The difference between these groups in their mean fear judgment scores was not statistically significant. However, a significantly greater proportion of the Fear-Suppression group in comparison to the Fear-Expression group used

<sup>4</sup>It is possible to conceive of a circumstance in which "suppression" of affect might lead to the absence of projection. Thus, if the S were completely successful in restricting his responses to those which were incompatible with fear responses, he might not experience any fear at all, and consequently he would not project since the present theory supposes that the response produced affective cues must be part of the stimulus field in order for projection to occur. The concept "denial" is perhaps more descriptive than are "suppression" or "inhibition" of the situation in which a particular affective response is elicited but the S is motivated not to recognize it as such.

extreme fear categories in describing the stimulus person.

3. The experimental differences are more striking for judgments of the fearfulness and anxiety of the stimulus person in situations other than that immediately depicted in the film.

4. The tendency to perceive the stimulus person as a fearful and anxious individual was significantly correlated with the tendency to perceive him as aggressive.

The data indicate that the arousal of fear results in a tendency to project fear onto a stimulus object in the environment. The results are also consistent with the hypothesis that suppression facilitates the tendency to project although, in this latter instance, one is less confident in rejecting the null hypothesis. Several alternative explanations of the effects

of the suppression variable were considered and the role of cognitive variables in the projection process was discussed.

#### REFERENCES

1. BELLAK, L. On the problems of the concept of projection. In L. E. Abt & L. Bellak (Eds.), *Projective psychology*. New York: Knopf, 1952. Pp. 7-32.
2. DOLLARD, J., & MILLER, N. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
3. FREUD, S., *Totem and taboo*. London: Routledge & Kegan Paul, 1950.
4. MURRAY, H. A. The effect of fear on the estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, 4, 310-329.
5. SEARS, R. R. Experimental studies of projection: I. Attribution of traits. *J. soc. Psychol.*, 1936, 7, 151-163.
6. WERNER, H., *The comparative psychology of mental development*. (Rev. ed.) Chicago: Follett, 1948.

Received December 4, 1956.



# THE MASCULINITY-FEMININITY DIMENSION IN NORMAL AND PSYCHOTIC SUBJECTS<sup>1</sup>

MAX R. REED<sup>2</sup>

*Washington University, St. Louis*

IN THE course of relating personality organization or structure to adjustment level or prognostic outcomes, the clinician must routinely attempt to assess the influence of the masculinity-femininity dimension. Psychoanalytic theory (3, 7, 8, 20) assigns a central role to this aspect of personal organization. Masculinity-femininity is viewed as a complex affair, derived from the combined effects of a number of personality determinants. Explicit relationships are postulated between genetic occurrences, body image or body feeling, and levels of ego organization (3, pp. 35-113, 418, 419). Early, severe disruptions in the sexual identification processes are thought to bring about major disturbances in body image organization (20, p. 252). The result is conceived as an admixture of cross-sex elements, a condition discordant with the body image essential to one's normal sexual role.

Such states are accounted for in psychoanalytic theory by postulating differences in levels of ego organization at the times at which early body image organization, primary identifications, and secondary identifications occur. Body image organization and primary identifications are thought to occur at a preverbal level. Fenichel (3, p. 107) suggests that primary identifications and the deepest layer of the ego are constructed through kinesthetic, olfactory, and visual sensations. On the other hand, secondary identifications and the superego are later developments, based primarily on auditory stimuli. Thus, there may be a wide discrepancy between the way a person knows that he should feel or act sexually and the way that he does feel or is capable of acting. These disparities are believed to be important contributors to a weakening of ego organization, the essential condition for psychotic illness.

The development of tests reflecting different levels of masculinity-femininity organization has paved the way for experimental investigation in this area. One line of development has made use of drawings of one type or another which, it is believed, reflect the basic masculinity-femininity organization. Franck and Rosen (6) devised a Drawing Completion Test (DC) which successfully differentiated college men and women. They reasoned that the differences in performance on this instrument reflected basic sex differences in body structure and psychosexual organization. The Draw-A-Person Test (DAP) is another drawing task believed to reflect sex differences in body image or body feeling and sexual organization. A number of studies (1, 4, 9, 11, 14, 15, 16, 22) with the DAP have yielded contradictory findings, but the test's full potential is by no means exhausted.

A second development has been the construction of verbal inventory measures which are thought to reflect the masculinity-femininity aspect of personality organization. Prominent among these measures is the MMPI Masculinity-femininity Scale (Mf), an empirically derived inventory yielding significant differences between men and women. Shepler (21) has shown that this test measures something different from the DC. While he found that both the Mf scale and the DC test differentiated men and women, he also found a low and insignificant correlation between the two test variables.

The DC, Mf, and the DAP were used in the present study to test the following hypotheses, derived from psychoanalytic theory, concerning the nature of differences in the masculinity-femininity organization of normal and psychotic women:

1. Psychotic women obtain more masculine scores on measures of body image than do normal women. (a) With the DC as a measure of body image, psychotic women yield more masculine scores than do normal women on

<sup>1</sup> This study is a condensation of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Washington University, St. Louis. The writer wishes to express his appreciation to Dr. Frederick H. Kanfer for his guidance in the study and for his critical evaluation in preparation of the present manuscript.

<sup>2</sup> Now at The Manatee-Sarasota County Guidance Center, Sarasota, Florida.

this measure. (b) With the DAP as another such measure, the relative body proportions of normal subjects' male and female drawings approximate more closely actual anatomical measures of such proportions than do the like body proportions of psychotic subjects' figure drawings.

2. The verbal statements of normal women concerning their masculinity-femininity status correspond more closely to their respective measures of body image status than do those of psychotic women. (a) Psychotic women show larger differences between their DC and Mf scores than do normal female subjects. (b) Psychotic women obtain higher (more masculine) DC scores than MMPI Mf scores, whereas normal female subjects obtain DC scores that are roughly equal to or lower (more feminine) than their corresponding Mf scores.

### METHOD

The study included an initial test of the hypotheses (designated Experiment I) and a replication study employing independently chosen samples and independent scorers (designated Experiment II). Experiment II was planned to examine any refinements of the original hypotheses which Experiment I might suggest and to cross validate the findings.

### Selection of Subjects

The control Ss for Experiment I were 22 nurses chosen randomly from one class of nurses in training at the St. Louis State Hospital. The control Ss for Experiment II were 28 nurses chosen in the same manner from the immediately following class of nurses at the same institution. The mean age for these Ss in Experiment I was 20.63 years with a standard deviation of 2.49. For Experiment II the mean age was 20.14 with a standard deviation of 1.19. All Ss were unmarried.

Twenty-five female psychotic Ss were selected for each experiment. The majority of these Ss for Experiment I was obtained from the St. Louis State Hospital, and for Experiment II a majority was drawn from the Elgin State Hospital at Elgin, Illinois. The psychotics were matched with the nurses on level of education, but it was not possible to match the two groups on the age variable. The mean age of the psychotic Ss of Experiment I was 29.28 years with a standard deviation of 3.21. In Experiment II, the mean age for the psychotic Ss was 28.52 years with a standard deviation of 3.54. In Experiment I, fifteen Ss were married and ten single. Sixteen Ss were married and nine Ss were single in Experiment II. In Experiment I, the Ss had been hospitalized from a minimum of one to a maximum of nine years with the mean length of hospitalization falling at 3.88 years. In

Experiment II, the length of hospitalization ranged from one to ten years with the mean at 3.96 years. The psychotic Ss in no case exhibited central nervous system involvement, nor had any S undergone shock therapy of any type within three months immediately preceding test administrations.

### Tests and Scoring Procedures

All Ss were administered the DC, the group form of the MMPI, and the usual DAP procedure.

*The MMPI Mf.* The published hand key was used to score the Mf scale of the MMPI for all Ss.

*The DC.* The DC protocols were all scored according to the criteria given in Franck's mimeographed manual (5). Although Franck and Rosen's (6) published scoring reliabilities on female normative groups are high, it seemed desirable to obtain interscorer reliabilities on the different populations used in the present study. When the writer and two additional scorers with graduate training in psychology scored the 25 psychotic records used in Experiment I, the resulting reliability coefficients were .79, .82, and .66. The writer's scoring was used for all statistical evaluations in Experiment I. In Experiment II, the independent scoring of one of the graduate students who participated in the first experiment was used.

In order to investigate Hypothesis 2 (b), DC standard scores were computed and the scale inverted for purposes of comparison with the MMPI Mf. Thus, for this comparison, a low DC standard score is in the feminine direction, whereas a low DC raw score is in the masculine direction.

*The DAP.* The following body measures were taken from each figure drawing protocol: overall height, waist height, neck height, above waist height, and arm length. On the basis of Dickinson's (2) findings that they reveal most clearly and consistently differences between males and females in actual measurements, the single measure of overall height and the ratios of arm length/waist height, above waist/waist height, and waist height/neck height were employed to test Hypothesis 1 (b). In these measures, the predicted proportional relationships for normals are as follows: (a) male figure drawing taller than corresponding female, (b) male proportions of arm length/waist height larger than the like female figure proportion, (c) female proportion of above waist/waist height smaller than the respective male figure proportion, and (d) female proportion of waist height/neck height greater than the respective male figure proportion. An additional summary measure consisted of the number of Ss for whom all of the differences in body proportions between the male and female figure drawings were in the normal anatomical direction.

As a check on the reliability of these measures, interscorer correlations were obtained for the DAP protocols of the psychotic Ss employed in Experiment I. Three scorers with graduate training in psychology measured all of these figure drawings. Product-moment correlations were calculated between the three sets of measures obtained. The resulting 30 reliability coefficients ranged from .91 to .99. Consequently, only the *E*'s measures were used in the statistical tests of hypotheses.



TABLE 1  
COMPARISON OF DC SCORES FOR NURSES AND PSYCHOTICS

Item	Experiment I		Experiment II	
	Nurses	Psychotics	Nurses	Psychotics
No. of Cases	22	25	28	25
Mean <sup>a</sup>	23.36	15.60	20.14	15.00
SD	3.41	2.70	3.13	2.59
Combined median	18.00		17.60	
No. above median	18	5	25	6
No. below median	7	17	3	19
Chi square	8.546		8.902	
<i>p</i>	<.005		<.005	

<sup>a</sup> Higher scores are in "feminine" direction.

### RESULTS

From Table 1 it is seen that the psychotic Ss in both experiments obtained significantly less feminine DC scores than the normal Ss. The nonparametric median test (18) yielded significant chi-square values ( $p < .005$ ) for the two comparisons. (All chi-square values obtained in this study were corrected for continuity.)

While these findings strongly support the hypotheses advanced, the possibility must be considered that factors other than the experimental variable contributed to the differences in DC scores. Since it was not possible to equate control and experimental groups for age, product-moment correlations were calculated between the DC scores and the Ss' ages. The values for the psychotic Ss were .20 and .04 for Experiments I and II respectively. For the nurses, the corresponding correlations were .42 and -.09. All correlations in the present study were tested for significance using the *r*-to-*z* transformation (17). None of these values are significant. Thus, there seems to be little appreciable influence of age on DC scores.

Evaluation of the relationship of marital status to DC performance was limited by the nature of the samples. In Experiment I, the 10 single psychotic Ss obtained a mean DC score of 13.7 with a standard deviation of 2.3, whereas the 15 married Ss obtained a mean of 16.8 and a standard deviation of 2.2. In Experiment II, the 9 single Ss obtained a mean DC score of 14.3 with a standard deviation of 3.1, and the 16 married Ss obtained a mean of 15.3 with a standard deviation of 2.1. For Experiment I the difference is significant ( $t =$

3.41,  $p < .01$ ), but for Experiment II the difference is nonsignificant ( $t = .79$ ,  $p > .10$ ). Although the absolute numerical means of the married Ss are more feminine than the DC means of the total psychotic groups, they are nevertheless significantly more masculine than the corresponding nurses' mean scores (Experiment I,  $t = 7.13$ ,  $p < .01$ ; Experiment II,  $t = 6.11$ ,  $p < .01$ ). Thus, in this study there is no conclusive evidence for a systematic relationship between DC scores and marital status.

Length of hospitalization similarly seemed to have no relationship to DC performance (Experiment I,  $r = -.12$ ,  $p > .05$ ; Experiment II,  $r = .30$ ,  $p > .05$ ). Thus, there is no evidence that extraneous factors account for the obtained differences in DC test performance.

Two additional factors must be considered in evaluating the DC score differences obtained by the experimental and control groups. First, the nature of the test suggests the possibility that simplified responses, a result of impoverished psychotic ideation, might be the major source of the more masculine scores obtained by the experimental groups. However, a rough classification of the drawings into simplified versus several other categories revealed that only 27 per cent of the psychotic drawings scored as masculine could be considered simplified. Since 25 per cent of the drawings scored as feminine could also be classed as simplified, there is little evidence that the more masculine scores of the experimental group resulted primarily from simplified drawings.

Second, in any group of Ss scoring similarly on the test, several different styles of drawings are found. This finding is in agreement with the results of Franck and Rosen (6). It is possible that such variations may be in some systematic way associated with different types of feminine organization of different adaptive values. Further work with the DC, aimed at refinements in the scoring system, should add to an understanding of the relationships of masculinity-femininity structure to levels of ego strength or adaptive capacity.

The DAP results, presented in Tables 2 and 3, give less convincing support for the first hypothesis. Only three of the ten chi-square values are significant.

TABLE 2  
GROUP COMPARISONS ON DAP MEASURES

Experiment I					
DAP Measure	Nurses		Psy- chotics		p
	F <sup>a</sup>	M <sup>b</sup>	F <sup>a</sup>	M <sup>b</sup>	
Height	13	9	7	16	2.669
Arm length/Waist height	11	1	10	10	4.088
Above waist/Waist height	18	4	12	12	3.811
Waist height/Neck height	21	1	15	8	4.355
Experiment II					
DAP Measure	Nurses		Psy- chotics		p
	F <sup>a</sup>	M <sup>b</sup>	F <sup>a</sup>	M <sup>b</sup>	
Height	16	12	8	14	2.706
Arm length/Waist height	12	3	9	4	0.427
Above waist/Waist height	20	2	12	7	3.118
Waist height/Neck height	21	1	13	6	2.814

<sup>a</sup> Subjects producing drawings yielding body proportions in normal anatomical relationships.

<sup>b</sup> Subjects producing drawings distorted by cross-sex proportions.

TABLE 3  
GROUP COMPARISONS ON COMBINED DAP MEASURES

Item	Experiment I		Experiment II	
	Nurses	Psy- chotics	Nurses	Psy- chotics
Number producing 100% body proportions in normal direction	11	4	14	4
Number producing less than 100% body proportions in normal direction	11	18	2	13
Chi square	3.642		5.430	
p	<.06		<.02	

Again, there was no systematic relationship between DAP measures and age. Product-moment correlations between the two variables ranged from  $-.44$  to  $.31$  with none of the values reaching significance. This finding is consistent with that of Lehner and Gunderson (14) that the figure drawings of women show no effects of age differences in the range from 20 to 40 years.

Inspection of Table 4 reveals no conclusive evidence for a systematic relationship between the DAP measures and other test scores. None of the correlations between the DAP measures and Mf scores are significant. Only one DAP measure yields a significant correlation with

TABLE 4  
TEST INTERCORRELATIONS

Measure	Test	Experiment I		Experiment II	
		Nur- ses	Psy- chotics	Nur- ses	Psy- chotics
Waist height	DC	.15	-.13	.00	-.56**
	Mf	.04	.17	-.10	.09
Arm length	DC	.37	-.65**	-.54	.10
	Mf	-.23	.00	-.11	-.29
Neck height	DC	.08	-.20	.12	-.50**
	Mf	.07	.21	.04	-.10
Above waist	DC	-.14	.16	.00	.56**
	Mf	-.02	-.12	.08	-.08
Height	DC	.07	-.96**	.04	-.42*
	Mf	.09	.18	.00	.32
Correlation between Mf and DC		-.21	.34	.07	.06

Note.—Only measurements of female drawings appear in table.

\*  $p < .05$ .

\*\*  $p < .01$ .

TABLE 5  
COMPARISON OF EXPERIMENTAL GROUPS ON DC-Mf STANDARD SCORES

Item	Experiment I		Experiment II	
	Above Md.	Below Md.	Above Md.	Below Md.
Nurses	17	5	18	10
Psychotics	7	18	8	16
Chi square <sup>a</sup>	8.546		4.950	
p	<.005		<.04	

<sup>a</sup> Chi square based on median test.

the DC scores in both experiments, and this only with the psychotic groups. Thus, there is considerable evidence in this study that the DAP is measuring something different from both the DC and Mf.

In agreement with Hypothesis 2 (a), the psychotic Ss of both experiments obtain larger absolute difference scores (Mf standard scores minus DC standard scores). Table 5 presents the significant chi-square values obtained when the median test was applied to these differences. However, some clarification of this finding is needed. The result is due primarily to the more feminine DC scores obtained by the normal Ss rather than to any differences in the Mf scores between the two groups. The Mf contributed little in discriminating the two groups.

A significantly larger number of psychotic than normal Ss obtained DC scores that were more masculine than their corresponding Mf



TABLE 6  
GROUP COMPARISONS ON MMPI Mf SCALE  
(Mf measures in raw scores)

	Experiment I		Experiment II	
	Nurses	Psychotics	Nurses	Psychotics
<i>N</i>	22	25	28	25
<i>M</i>	36.59	35.76	36.10	35.50
<i>SD</i>	4.69	5.73	6.82	4.80
<i>t</i>	1.63		1.045	
<i>p</i>	>.10		>.30	

scores, giving strong support for Hypothesis 2 (b). Chi-square values for differences obtained in Experiments I and II, respectively, are 14.41,  $p < .001$ , and 4.05,  $p < .05$ .

From Table 6 it can be seen that the Mf alone, in agreement with the findings of Renaud (19), is of little value in discriminating between major diagnostic groups differing greatly in sexual organization, and contributes little toward differentiating the normal and psychotic groups. The findings of Gough (10) that psychotic Ss obtain more variable scores on the Mf scale than do normal Ss are not borne out. Product-moment correlations calculated between Mf and DC test scores give further evidence, in agreement with Shepler (21), that these two instruments assess different attributes. For the normal Ss, the correlations were  $-.21$  and  $.07$  for Experiments I and II respectively. For the psychotic Ss, the corresponding correlations were  $.34$  and  $.06$ . For samples of this size, these values may not be considered as different from zero.

In order to evaluate better the relative efficiency of the two types of measures (DC scores and difference scores) in differentiating the criterion groups, the data were combined from the two experiments, and optimal cutting points were determined for the distributions of DC and difference scores. Table 7 shows that the DC scores alone discriminated much more effectively. The optimal DC cutting score of 17.5 misplaced 18 Ss. Seven nurses (14 per cent) fell on the psychotic side of the distribution. Eleven psychotic Ss, 22 per cent of the total psychotic group, fell in the nurses' end of the distribution. The optimal difference score (Mf standard score minus DC standard score) of  $+5.5$  misplaced 33 Ss. Twenty-four nurses, 48 per cent, fell at the psychotic end of the scale while nine, or 18 per cent, of the psychotic Ss were misplaced. The correct selections for both

TABLE 7  
DISTRIBUTION OF DC AND DIFFERENCE SCORES ABOUT  
OPTIMAL CUTTING POINTS, COMBINING DATA FROM  
EXPERIMENTS I AND II

Item	DC Scores		Difference Scores	
	Above 17.5	Below 17.5	Above +5.5	Below +5.5
Nurses	43	7	26	24
Psychotics	11	39	9	41
Chi square	39.668		12.362	
<i>p</i>	<.001		<.001	

tests are significantly beyond chance expectancy ( $p < .001$  in both instances).

A word of caution is added here. Although some of the differences obtained between criterion groups in the present study are highly significant, considerable overlap occurs in all cases. Even though the two experiments were conducted on independently chosen groups, the samples were small and some drop is evident in the significance levels obtained in the second experiment. More extensive and refined investigation needs to be carried out in this area before these measures can be employed diagnostically or for purposes of individual prediction.

# DISCUSSION

The results of this study are interpreted as being in general agreement with hypotheses derived from psychoanalytic theory (3, 7, 8, 20) concerning the nature of differences along a masculinity-femininity dimension in normal and psychotic women.

The evidence supports the hypothesis that there exist two rather clearly defined levels of functioning within this dimension. One level seems to be associated with or bound to kinesthetic or perceptual-motor facets of expression and can be thought of as falling within the psychoanalytic id structure. The second level of organization appears to be more closely associated with verbal expression, and thus seems more closely associated with cultural expectations of the superego. For the psychotic groups in this study, disturbances at the nonverbal level of organization are crucial and are indicated by markedly masculine performances on the DC. It is suggested that such a deviant organization provides a core of unacceptable impulses and a steady source of life stress which contribute to the psychotic Ss'

problems of adaptation and to their eventual breakdown. Extensive research is needed to lay the groundwork for a more refined and comprehensive theory concerning this aspect of human development.

Future research must also determine whether the differences in DC scores represent differences in developments in the identification process or reflect some basic somatic differences, such as differences in hormonal function, or both. In addition, the relationship of styles of sexual adjustment to general adaptive functioning, and the relationship between masculinity-femininity organization and psychosis in male populations need to be explored.

### SUMMARY

The present study was designed to test hypotheses derived from psychoanalytic theory that psychotic and normal women differ in their responses to measures of different levels of masculinity-femininity organization. The DC and the DAP were employed as nonverbal measures of body image status, and the MMPI Mf Scale was administered as a verbal measure. On the DC, the psychotic women obtain more masculine scores than do normals. However, using the DAP as a measure of body image, only two of four measures yielded significant differences, and these did not hold up on cross validation. The verbal statements of normal women concerning masculinity-femininity status correspond more closely to their respective measures of body image status than do those of psychotic women.

A word of caution was added concerning the DC. Although some of the differences obtained with the test between criterion groups were highly significant, the amount of overlap was too great to allow its use for individual prediction.

Theoretical implications of the study were discussed, and areas for future research indicated.

### REFERENCES

1. BARKER, A. J., MATHIS, J. K., & POWERS, C. A. Drawing characteristics of male homosexuals. *J. clin. Psychol.*, 1953, **9**, 185-188.
2. DICKINSON, R. L. *Atlas of human sex anatomy*. Baltimore: Williams & Wilkins, 1949.
3. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: W. W. Norton, 1945.
4. FISHER, S., & FISHER, RHODA L. Style of sexual adjustment in disturbed women and its expression in figure drawings. *J. Psychol.*, 1952, **34**, 169-179.
5. FRANCK, KATE. *Mimeographed manual for the Drawing Completion Test*. 1949.
6. FRANCK, KATE, & ROSEN, E. A projective test of masculinity-femininity. *J. consult. Psychol.*, 1949, **13**, 247-256.
7. FREUD, S. The passing of the Oedipus Complex. *Collected papers*, Vol. II. London: Hogarth Press, 1942.
8. FREUD, S. *The ego and the id*. London: Hogarth Press, 1947.
9. GOLDWORTH, S. A comparative study of the drawings of a man and a woman done by normal, neurotic, schizophrenic, and brain-damaged individuals. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1950.
10. GOUGH, H. G. Diagnostic patterns on the MMPI. *J. clin. Psychol.* 1946, **2**, 23-37.
11. GRONICK, S., & SMITH, L. J. Sex sequence in the Draw-A-Person Test and its relation to the MMPI Masculinity-femininity Scale. *J. consult. Psychol.*, 1953, **17**, 71-73.
12. HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. (Rev. ed.) Minneapolis: Univer. of Minnesota Press, 1944.
13. JOURARD, S. M., & SECORD, P. F. Body cathexis and the ideal female figure. *J. abnorm. soc. Psychol.*, 1955, **50**, 243-246.
14. LEHNER, G. F. J., & GUNDERSON, E. K. Height relationships on the Draw-A-Person Test. *J. Pers.*, 1953, **21**, 392-399.
15. MACHOVER, KAREN. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.
16. MAINFORD, FLORENCE F. A note on the use of figure drawings in the diagnosis of sexual inversion. *J. clin. Psychol.*, 1953, **9**, 188-189.
17. MCNEMAR, Q. *Psychological statistics*. New York: John Wiley & Sons, 1955.
18. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122-143.
19. RENAUD, H. R. Clinical correlates of the masculinity-femininity scale of the Minnesota Multiphasic Personality Inventory. Unpublished doctoral dissertation, Univer. of California, 1950.
20. SCHILDER, P. *The image and appearance of the human body*. New York: International Universities Press, 1950.
21. SHEPLER, B. F. A comparison of masculinity-femininity measures. *J. consult. Psychol.*, 1951, **15**, 484-486.
22. SMITH, ELGIE. A study of sex differentiation in drawings and verbalizations of schizophrenics. *J. clin. Psychol.*, 1943, **9**, 183-185.

Received June 8, 1957.



# GALVANIC SKIN RESPONSE AS A MEASURE OF PATIENT'S REACTION TO THERAPIST'S PERMISSIVENESS<sup>1</sup>

JAMES E. DITTES

*Yale University*

**M**ost theories of the dynamics of psychotherapy agree on the importance of an attitude manifested by the therapist toward the patient, which is variously characterized as accepting, warm, nonthreatening, or permissive. But it is not always made clear what types of behavior define this "acceptance" or what is its function within the situation of psychotherapy.

"Acceptance" or "permissiveness" may be defined as an absence of various forms of punishing behavior which a patient is accustomed to receive from other persons. The nature of such anticipated punishments differs among individuals but frequently includes such cues as these: other persons fail to pay close attention to one's words; others seem unwilling or unable to understand one's own point of view or remarks; one has difficulty making his points understood; not only do others fail to give a fair hearing, but they also insist vigorously on their own point of view, interrupting, contradicting, disregarding and not attempting to reconcile disagreement; they merely announce their position, as though one's own view isn't even worth considering. Such types of behavior are punishing, in large part, because they communicate to a person that he is regarded with low esteem and as of negligible importance. Less subtle forms of such punishment are outright criticism, rebuke, expressions of rejection, and clear indications that one is considered inadequate in some respect and should change. Sympathy and friendliness may be expressed, but in a patently forced and unconvincing manner which says to a person that he is of little enough importance that he can be dealt with in a manipulative, contrived manner and not as an equal.

<sup>1</sup> This study is part of a psychotherapy research program supported by United States Public Health Service Grant M-648 and supervised by Dr. John Dollard. The data used here had been previously obtained by Dr. Dollard and Dr. Frank Auld, Jr. Appreciation is due Dr. Auld, Dr. Dollard, and Dr. George Mahl for making suggestions on the manuscript and to Dr. Auld, Mrs. Ann Byler, and Mrs. Michele Morris for serving as judges in ratings of the verbal material.

Such types of behavior are functionally equivalent whether they are experienced as threats of punishment or as painful punishment in themselves, since such experiences become either threatening or painful as they become associated with clearly painful experiences, such as extreme rejection and withdrawal of important satisfactions; and whether as punishment or as threat of punishment, such cues motivate avoidance behavior. For example, the stimulus of having one's words not attended to carefully may frequently be associated with the frustration of not receiving what has been requested or with such clearly painful experiences as being ignored or otherwise denied signs of regard or affection by the same person who has half-heartedly listened to one's words. Such a cue comes to motivate avoidance whether it is regarded as itself punishing or merely as a signal threatening punishment.

To such punishment or threats of punishment, the initial response of a person is likely to be some state of reaction which may be characterized as tension, anxiety, alertness, mobilization, a state which undoubtedly has a variety of physiological components, and which comes to be followed by, and thereby lessened by, some type of behavior which serves to escape or otherwise defend against the punishment. Within small groups, for example, it has been demonstrated that cues indicating low acceptance of a person by other group members results in increased conformity by that person to norms of the group, even in the face of evidence contradicting the norms (3). Among the types of behavior which are learned in response to such threats of socially administered punishment are probably those which are of interest to a therapist as neurotic symptoms and as defensive resistance within therapy.

In therapy, "permissiveness"—the absence of such punishments and threats in the behavior of the therapist—should serve two related functions (4). First, the absence of such threatening cues should allow the patient to speak and act without eliciting in him the stress or tension and accompanying defensive avoidance

measures which are ordinarily evoked by such cues in the behavior of others. Second, whenever either the therapist's behavior or the patient's own thoughts or other behavior actually do produce cues which have been accompanied by punishment in the past and hence serve as threats, the failure of these to be reinforced by actual punishment from the therapist should serve to extinguish the effectiveness of the cue as a threat: for example, the failure to be heard clearly will no longer elicit a state of tense readiness followed by some defensive measure; it will be found not to be a signal of punishment.

In practice, it is unlikely that any behavior of the therapist, even the most permissive, completely avoids cues which are threatening to the patient. As a protective, adaptive measure, a neurotic patient has probably learned to respond to a wide variety of cues as though they were threats and has acquired a "low threshold" of perception and response to such cues.

The resistive or defensive responses elicited by any such cues during therapy are probably difficult to detect and to measure for several reasons. There is much individual difference in the forms of such reactions. Any individual has a large repertory of such responses, and very different ones may be elicited by slightly different situations. Because many forms of resistance and defense themselves are frequently punished, particularly the most overt, the patient learns to be very subtle in the resistant response; the resistant, avoidant behavior—particularly of a well-practiced defender, as most patients are—probably deviates only very slightly, by the minimum possible, from ordinary nonresistive behavior, so that the resistance itself is most likely to escape detection and consequent punishment.

A potentially more stable indicator of a patient's perception and response to cues of threat is the physiological measure of the galvanic skin response. Probably alone among physiological responses, the GSR is activated solely by the sympathetic and not the parasympathetic system and is generally characterized as indicating a state of "mobilization" or "preparation" (1, 7). It seems plausible to suppose that the GSR reflects the general tension or "anxiety" or state of readiness which is the immediate reaction to a threatening cue.

The GSR record of a patient in therapy, then, should show a relationship with the "permissiveness" of the therapist, as defined in terms of the absence of various punishments or cues threatening punishment, and measures of resistance should show an increase with decreased permissiveness and increased GSR.

## METHOD

This study is based on the records of the final 43 hours of therapy with a 36-year-old, white, male, neurotic patient, seen by an experienced therapist three or four times a week. The patient previously had 30 weekly sessions with the same therapist before therapy was recessed for the summer. The patient terminated therapy at the end of the series reported here. All sessions were electrically recorded and verbatim transcripts typed. A continuous record of the patient's palmar skin conductance was made during each session with apparatus previously described (2). In one session, the 14th of the present series, recording equipment failed, and this is excluded from the study.

### *Estimation of Therapist's Permissiveness*

Using the typed transcripts, two judges independently evaluated the remarks of the therapist, using ratings of one to four on each of four scales intended to represent aspects of permissiveness. Hours were judged in random order. The four scales were defined as follows:

*Attentiveness.* Patient is able to feel self continually in the therapist's thoughts. There is evidence that the therapist is listening carefully and remembering what the patient tells him, that he is not preoccupied or distracted.

*Understanding.* The therapist shows that he understands the patient's thoughts and feelings as the patient means and feels them. The patient has no trouble getting his points across; the therapist is not stupid or obtuse. The therapist's interventions suggest thoughts or feelings which the patient can recognize as his own and do not seem to him farfetched.

*Gentleness.* Therapist fails to show any of the signs of rejection or disapproval which the patient may have come to expect from others. He does not condemn, criticize, scold, judge, belittle, or use sarcasm or irony. He does not give evidence of a feeling of the patient's inferiority by urging the patient to change some present behavior, by interrupting or trying to channel the patient's speech, or by insisting on his own interpretations either openly or subtly; he speaks tentatively and offers his notions only as a possibility for the patient to consider.



*Over-all acceptance.* The total impact of the therapist's activity is likely to make the patient feel that he is unconditionally accepted, respected, valued, "loved," is an important person, not under judgment or threat of potential punishment, is "safe" to venture thoughts and feelings which have been punished in his previous relations with other people.

Judgments on each scale were made for the total hour and for smaller units of the therapist's remarks within each hour. A unit was defined as all the remarks of the therapist separated on the typescript by less than a page of the patient's remarks. The therapist units, in other words, were separated from each other by a page or more of the patient's speech. Most hours had from four to six such units. Agreement between the two judges in judging the whole hours with the four scales summed yielded a correlation of .69<sup>2</sup>, and the correlation of judgments in the smaller units within hours was .61. Four principal scores were used.

*Whole hour permissiveness* is the sum for the four scales of judgments of the hour as a whole, with both judges' scores summed. Scores ranged from 15 to 32 with the median and mean at 21. This score was used as the measure of permissiveness in all results reported unless some other score was specifically indicated.

*Initial permissiveness* was measured by averaging the ratings on all therapist units in the first five pages of typescript.

*Opening friendliness.* The behavior of the therapist in greeting the patient and in any interaction during application of the GSR electrodes was judged by two judges with respect to "friendliness," defined as responsiveness to incidental social overtures of the patient or as initiation of such conversation. Independent ranking of 36 hours by two judges yielded a Spearman rank-order correlation of .87. The six remaining hours were those both judges eliminated from the sorting as providing insufficient basis for judging; in most of these, only a bare greeting was recorded. The average rank of the two judges was used as the score for the hour on this measure.

*Change in permissiveness.* Instances of marked change in permissiveness within any hour were defined as any change of four or more points (representing an average of one or more points for each of the four scales) between two succes-

sive therapist units. Nineteen such instances occurred in the 42 hours.

### *Scoring of GSR Record*

A GSR was scored whenever the continuous record of conductance met one of the following criteria: (a) a rise at the rate of 4000 or more ohms per second with an amplitude of at least 6500 ohms; or (b) a rise of at least 10,000 ohms amplitude at a rate of 1333 or more ohms per second. Such a change was of the order of 3 per cent of the basal resistance level. These criteria had been used previously (2) and were found to be easily distinguishable with the calibrated paper used; and had been used to distinguish between clear-cut abrupt rises in conductance and much smaller fluctuations.

The GSR measure primarily used in this study was the frequency of such rises for each 50-minute therapy session. This frequency was approximately normally distributed with a mean of 45 and median at 43, ranging from 11 to 90. In a randomly selected set of half the hours, this frequency score correlated .95 with the total amplitude of GSR deflections within each hour. In the same sample of hours, the GSR frequency correlated -.42 with the amount of decrease during the hour of the basal resistance level, as calculated by dividing the amount of change in ohms by the beginning resistance in ohms.

For obtaining a GSR score within any part of an hour, the rate of GSRs per minute was used. This varied between .25 and 3.2 GSRs per minute in the hours examined. The criterion for determining whether a *change in rate* occurred was a difference of .25 or more GSRs per minute between successive portions of an hour.

To measure GSR rate following any therapist intervention, GSRs were counted from five minutes after the beginning of the therapist unit until five minutes after the beginning of the next unit, and the rate per minute computed. It was reasoned that the effect of the therapist's speech on the GSR would appear only after the patient had heard a substantial amount of the unit. The time of five minutes was arbitrarily selected before the data were examined. The unit of the therapist's speech typically covered somewhat more than this period.

To measure the GSR preceding a therapist

<sup>2</sup> All correlation coefficients reported in this paper are product-moment correlations except as otherwise noted.

unit, GSRs were counted from the beginning of the preceding therapist unit to the beginning of the relevant therapist unit, and the rate per minute computed. It was assumed that this period best represents the unit of patient behavior which might be suspected of affecting the behavior of the therapist in the subsequent unit.

### *Measures of Patient's Behavior*

**Embarrassing sex statements.** Two judges, working independently from the transcripts, identified any statement by the patient clearly acknowledging personal sexual desires and behavior. Reliability for these judgments by these judges has previously been shown to be represented by phi coefficients of the order of .80 (2). A statement was scored as an embarrassing sex statement when selected by both judges.

**Resistance.** Two judges independently rated the patient's remarks as to whether they dealt with emotionally significant material or were resistive in the form of discussing less emotionally significant material, such as mechanics of therapy, trivial daily activity without emotional involvement, or another person's problems. The purpose of this measure was to check the possibility that a friendly, accepting manner by the therapist could be related to a low GSR score because it encouraged the patient to behave more as though he were in a friendly social conversation and hence discuss material less likely to involve emotions and the GSR. Ratings from 1 to 4 were made for each third of an hour and for the whole hour. The sum of these four judgments by both judges was taken as the resistance score. This score ranged between 8 and 24 with a mean of 17 and median at 18. Agreement between two judges yielded a correlation coefficient of .40.

**Interruptions in patient's speech.** As another possible indication of the kind of resistance which might be encouraged by a friendly therapist, a count was made of the number of times in each hour that the patient paused in his speech for five seconds or more. It was assumed that more fluent, less interrupted speech might accompany more sociable and less emotionally significant remarks and that more frequent pauses might accompany more significant material. Although using a slightly different indication of fluency and silence, Mahl

(5) has demonstrated that increased silence is associated with increased anxiety and that decreased silence is apparently an indication that other resistances than silence are operating. In the present study, the correlation of this score, number of pauses, with the judgment of resistance described above is  $-.29$  ( $p = .06$ ), i.e., higher resistance is associated with fewer pauses.

The frequency of pauses was also used in the analysis as a possible indication of depression in the patient.

## RESULTS

### *Relation between Permissiveness and GSR*

Each of the indicators of therapist's permissiveness showed a significant inverse relationship with frequency of galvanic skin response. Table 1 reports product-moment correlations between the measures.

Since it was also found that the GSR is related to the emotional significance of the patient's speech, as is reported in more detail below, this relationship partially obscures the relationship between therapist's permissiveness and GSR. If the emotional significance of the patient's remarks is made more homogeneous by eliminating the hours most extreme in resistance and those most extreme in emotional significance of content, the relationship between permissiveness and GSR becomes sharper. The last two columns of Table 1 report the correlation between permissiveness and GSR, disregarding, in one column, the hours which represent the top and bottom eighths on the distribution of resistance scores, and in the other column, the hours which represent the top and bottom eighths on the distribution of pause scores.

Changes in permissiveness from one hour to

TABLE 1  
RELATIONSHIP BETWEEN PERMISSIVENESS AND GSR

Measure of Permissiveness	Product-moment Correlations*		
	All Hours	Excluding Extreme Hours on	
		Resistance	Pauses
Whole hour permissiveness	-.51	-.65	-.73
Initial permissiveness	-.41	-.44	-.49
Opening friendliness	-.52	-.58	-.59

\* All relationships are significant at beyond the .01 level.



the next were accompanied by expected changes in the frequency of GSR. Fourteen hours showed a decrease of two or more points in permissiveness from the next previous hour. Nine of these showed an increase in frequency of GSR, four a decrease, and one no change. Of 14 hours showing an increase in permissiveness of two or more points, three increased in GSR frequency and 11 decreased. This relationship yields a chi square of 7.0 ( $p < .01$ ). This analysis disregards the three hours which, because of vacation or illness, followed the previous hour by a lapse of a week or more.

Eighteen embarrassing sex statements occurred, distributed among eight of the hours. A significant tendency was found for the ESS to be accompanied by a GSR in hours scored low on permissiveness and not to be accompanied by a GSR in more permissive hours.

Of 10 ESSs in four hours above the median on whole hour permissiveness, four are accompanied by a GSR and six are not. Of eight ESSs in hours below the median on permissiveness, seven are accompanied by GSR and one is not. This yields a chi square of 4.1 ( $p < .05$ ). If the ESSs are sorted at the median according to the permissiveness score of the therapist unit immediately preceding the ESS, the relationship shows a chi square of 5.8 ( $p < .02$ ). If opening friendliness is used as the measure of permissiveness, the relationship yields a chi square of 7.1 ( $p < .01$ ).

Although there is evidence that judges agree on the implicit unitizing in scoring ESSs (2), the unit of an ESS is not explicitly defined, and there accordingly exists the possibility of bias in determining whether or not a GSR accompanies a particular ESS. Therefore, an alternative analysis was made, using as an objectively defined unit of the patient's speech all his remarks occurring between pauses of five seconds or more or between remarks by the therapist. The rate of GSR per minute during speeches in which an ESS occurred was calculated. The average rate of GSRs for all the periods of ESS speech in hours below the median in permissiveness is 1.8 and for hours above the median is .8. The difference in rates between the two sets of ESS speeches is shown by a  $t$  test to be significant beyond the .05 level.

The four subscales originally intended to define permissiveness proved not to contribute equally to the relationship with GSR. The

TABLE 2  
RELIABILITIES AND INTERCORRELATIONS OF FOUR  
PERMISSIVENESS SUBSCALES WITH REGRESSION  
WEIGHTS FOR MULTIPLE CORRELATION  
WITH GSR

Subscale	Correlation Between Judges	Intercorrelations			Correlation With GSR	Regression Weight With GSR
		Und.	Gen.	Acc.		
Attentiveness	.14	.57	.35	.50	-.38	.24
Understanding	.53		.48	.57	-.32	-.03
Gentleness	.51			.85	-.51	.42
Acceptance	.50				-.48	.02

multiple correlation between the four scales and GSR is  $-.55$ . The interjudge reliability for each scale, the intercorrelations, and the regression weights are shown in Table 2. The gentleness scale alone correlates as well with the GSR ( $r = -.51$ ) as does the sum of the four scales. Apparently, understanding and over-all acceptance are not related to GSR except as they comprise the factor which may be labeled as gentleness or a nonpunishing manner, the chief component of permissiveness with which the GSR is related.

#### Alternative Interpretations

The general interpretation of these results is that the degree of apparent permissiveness or threat in the behavior of the therapist is responded to by the patient in a way which may be measured by the GSR. Before this interpretation is discussed in more detail, several alternative interpretations will be considered with additional results reported as they pertain to each one.

It could be contended that the GSR and the therapist's manner are both consequences of some aspect of the patient's behavior. For example, it may be suspected that on occasions when the patient is depressed, the GSR reflects this fact with a lower frequency and that the therapist also responds to this depression with a more kindly, sympathetic manner, which is judged more permissive. Or perhaps the patient comes to certain hours with a particular anxiety which results in more GSRs and which also results in a resistive kind of behavior with which the therapist attempts to deal by using more radical and direct interventions, which are judged less permissive.

There is evidence, however, that variations of permissiveness within hours are followed

Bureau Ednl. Sy. Res. Coll.  
DAVID HASE TRAINING COLLEGE

Dated.....

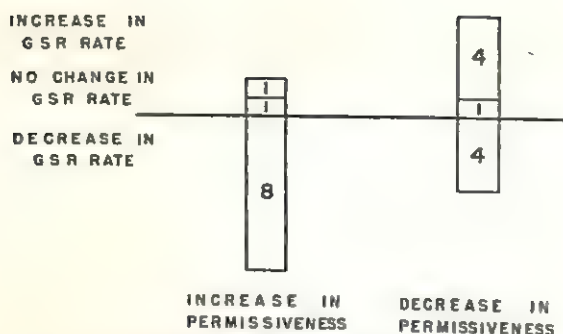


FIG. 1. EFFECT OF CHANGES IN THERAPIST'S PERMISSIVENESS ON SUBSEQUENT GSR RATE

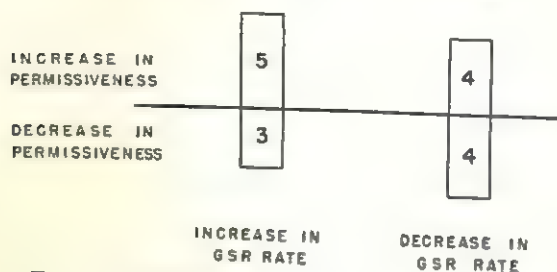


FIG. 2. EFFECT OF CHANGES IN GSR RATE ON THERAPIST'S SUBSEQUENT PERMISSIVENESS

Omitted are approximately 60 instances in which a change in the GSR rate is followed by no change in permissiveness.

but *not* preceded by corresponding inverse variations of the GSR. These data are reported in Figures 1 and 2, which record all within-hour changes of four or more points in permissiveness between two successive therapist units. The tendency of a change in permissiveness to be followed by a change in rate of GSR in the direction predicted by the interpretation held here yields a chi square of 2.7 ( $p = .05$ , one tail). But there is no relationship between changes in permissiveness and changes in the preceding GSR rate. These results support the interpretation that the GSR is in fact a response to the therapist's behavior.

With reference to the more specific possibilities that either depression or resistance of the patient affects both GSR and permissiveness, a possible measure of depression, frequency of pauses, though correlated with permissiveness ( $r = .38$ ,  $p < .02$ ) is not correlated with frequency of GSR ( $r = .14$ ). Resistance as judged in this study is correlated inversely with permissiveness ( $r = -.53$ ,  $p < .01$ ) and only slightly with frequency of GSR ( $r = .22$ ). The frequency of pauses score just reported may also be considered as an indication of resistance.

Another type of interpretation would explain the relationship of permissiveness and GSR as an artifact. It might be suspected that the therapist was judged more permissive and friendly when he talked more, thus leaving the patient less time to think and speak his own thoughts, and that listening is less likely to produce GSRs than is speaking. However, a count of the frequency of the therapist's speech (number of times therapist spoke between utterances of the patient) shows a negative correlation with permissiveness ( $r = -.33$ ,  $p < .05$ ) and no relation with GSR frequency ( $r = .07$ ).

Perhaps the most likely alternative explanation is that permissiveness, as defined here, encourages certain kinds of resistance by making the patient feel that the therapy is more like a social occasion, leading him to speak of routine, less emotionally involved material; hence, fewer GSRs are to be expected. Three types of evidence are available on this point.

First, this type of resistance was judged with moderate reliability between two judges but showed little relation with GSR ( $r = .22$ ), and furthermore yielded a negative relationship with permissiveness ( $r = -.53$ ) contrary to that proposed by the explanation.

Second, the other measure of resistance, number of pauses by the patient, was also found related to permissiveness ( $r = .38$ ,  $p < .02$ ) but not to GSR ( $r = .14$ ).

Third, variability in the patient's expression of emotionally significant material was presumably held constant in the judging of embarrassing sex statements, and as reported above, a clear relationship between permissiveness and GSR remains when such a uniform type of expression is examined. Of further suggestive importance is the fact that ESSs occur as frequently in permissive as in nonpermissive hours, contrary to what would be expected by the alternative explanation. ESSs did not, however, occur in hours of moderate permissiveness. All of the ESSs occurred in hours which rated either in the upper or lower thirds on the distribution of permissiveness.

#### *Relation of Resistance with Permissiveness and GSR*

Both measures of resistance show that in hours of moderate permissiveness (permissive-



TABLE 3  
RELATION OF RESISTANCE AND GSR AT DIFFERENT  
LEVELS OF PERMISSIVENESS

Degree of Permissiveness	Number of Hours	Correlation of GSR with	
		Judged Resistance <sup>a</sup>	Frequency of Pauses
High	12	.02	.33
Moderate	15	.48	.67
Low	15	-.15	.16
All hours	42	-.22	.14

<sup>a</sup> Signs of the resistance judgments are reversed so that, as with the pause score, a high score represents more emotionally significant, less resistive, content.

ness scores within two points of the mean), there is a clear relationship between frequency of GSR and the extent to which the patient's speech comprises emotionally significant, non-resistive content. These correlations are reported in the second line of Table 3. Within this same group of hours, the correlation of GSR and permissiveness is slightly positive ( $r = .16$ ); presumably the over-all trend of an inverse relationship between GSR and permissiveness is obscured by the strong relationship of GSR and content of patient's speech in these hours. The relation of GSR and emotionally significant content disappears when permissiveness becomes more extreme in either direction and presumably becomes the stronger determinant of the GSR as shown in the first and third lines of Table 3.

Each measure of resistance shows a significant correlation with permissiveness over all the hours. Correlation of permissiveness with judged resistance is  $-.53$  and with pauses is  $.38$ , in each case indicating more emotionally significant, less resistive content accompanying higher permissiveness.

#### DISCUSSION

The behavior of a therapist who preserves, in general, a nonpunishing, warm, accepting relationship with a patient may, nevertheless, include detectable variability in this characteristic, even within a narrow range, and offers to the patient at least occasional cues which he may regard as threatening or punishing. Such cues would include any behavior which may, correctly or incorrectly, communicate to the patient that the therapist regards him as an inferior, inadequate, or otherwise unworthy and unrespected person. Apparent failure by the therapist to understand the patient's words

or his interruption of the patient, or some form of scolding, or his apparent determination to force his own interpretations and points of view on the patient may be among such cues.

It appears confirmed by this study that a patient may react to such cues with a response which includes the galvanic skin response and which may represent a general tenseness and a defensive readiness to cope with the danger.

An episode occurring in one hour provides illustration and strong support for these interpretations. Hour 46, the contents of which were unknown to the investigator until near the end of the study, marks a distinctive change in the frequencies of GSR. Of the 14 preceding hours, only three have GSR frequencies under 50. Of the 28 following, only four have GSR frequencies over 50. A rank correlation ( $\rho = .49$ ,  $p < .01$ ) was found between sequence and GSR frequency, but this is due almost entirely to the change at this hour. Rank correlation for hours 46 and after is  $-.01$  and for hours before 46 is  $.23$ . (There is a slight correlation of  $.24$  between permissiveness and sequence, but this is not significant, and there is no corresponding sharp difference at hour 46.)

About one-third of hour 46 is devoted to interpretations by the therapist concerning his apparent friendliness to the patient and to responses by the patient on this subject. In response to the patient's observation that he can't communicate as freely as he consciously wants to, the therapist remarks, "I think you have to like somebody if you want to communicate. You have to feel free. . . Who's willing to confide or tell their private thoughts to other than a friendly person?" The patient, in effect, confirms the implications of the interpretation by reporting that he has felt the therapist to be brusque and unfriendly, as evidenced by his failure to call the patient by his first name and his reluctance to engage in social banter before and after the therapy sessions. He reports his feelings during an episode in the previous day's therapy when, "I was so anxious or falling all over myself. . . trying to impress you. . . as though I were rushing in to fill a gap that seemed to be necessary to fill to maintain a relationship, for instance. . . I say that type of thing often where I feel that I got to rush in and patch up a rupture rather than widen the rupture or act natural or expose myself."

These words of the patient seem to support the theoretical interpretation offered in this paper. In speaking of the "relationship," he presumably refers to various cues representing the therapist's regard and esteem for him, cues on which he is highly dependent. He indicates that the absence of such cues becomes a painful experience and motivates frantic efforts to defend against such a possibility. Such cues as not being called by his first name in spite of a long acquaintance and not finding a friendly response to efforts to initiate light social conversation apparently have been responded to by the patient as though they were indications of absence of esteem or some other punishing attitude, as in fact they normally would be in most social experience, and this response has included high GSR activity. Apparently, the effect of the therapist's interpretation and the patient's reply has been to let the patient realize that the therapist, unlike most other persons in his social experience, may be impersonal without being hostile or punishing. These particular cues in the therapist's behavior, though they continue to exist, lose their effectiveness as threats, and the GSR activity immediately drops and varies around a lower level.

The preliminary attempt to separate different aspects of permissiveness on separate scales results in the finding that absence of certain types of social punishment, defined here as "gentleness," proved to have the most significant relation with the GSR. In view of the unique and special emphasis sometimes given to "understanding" as the expression of a therapist's "acceptance" (6), it is of some interest to note that, as judged here and for this patient, this aspect of permissiveness is not related to the GSR independently of the scale which is most explicitly defined as the absence of punishment.

The judgment of resistance in this study was made with only moderate reliability and correlates only moderately with the objective measure of resistance, pause frequency. However, the two independent measures are highly consistent in their relationship with GSR frequency and with permissiveness and yield results that corroborate the theoretical interpretations proposed above.

If certain cues in the therapist's behavior are responded to with anxiety and fear of

punishment, it would be expected that such anxiety or fear would be further increased when the patient's own behavior or speech includes characteristics that may seem more "punishable" than when his behavior or speech is more "innocent." The combination of some cues in his own behavior that have been frequently followed by punishment plus cues in the therapist's behavior that threaten punishment should evoke more anxiety (and GSR) than the latter cues alone. Personal behavior which may have been punished in some way by others would include the discussion of sexual, hostile, and other feelings toward intimately related persons—behavior, in other words, which was judged in this study as emotionally significant, less resistive.

Within the limited range of moderate permissiveness, this is precisely the relation which was found: the emotional significance of the patient's speech is significantly correlated ( $r = .48$  or  $.67$ , depending on which measure is used) with the GSR, and is a more important determinant of the GSR than is permissiveness within these hours.

At lower levels of permissiveness, however, the findings are different. GSR is high and resistance is high (both being clearly correlated negatively with permissiveness), but the GSR and resistance are not correlated. These results suggest that anxiety (and GSR) are so uniformly high in response to the low permissiveness that they evoke generally resistive, nonintimate speech by the patient. The GSR is attributable chiefly to the low permissiveness, and not to emotionally significant content, because the latter is not present.

At high levels of permissiveness, GSR and resistance are both low, and again not significantly related with each other. Apparently, the permissiveness is great enough that the patient speaks freely of emotionally important material and does so without feeling under any threat. The cues of discussing emotionally vital matters may evoke GSR only as these cues enter into a kind of multiplicative relationship with cues of threatening behavior by the therapist. The results obtained with the ESS judgments further support these interpretations.

#### SUMMARY

Frequency of a patient's galvanic skin response is shown to be inversely related to the



judged permissiveness of the therapist during a series of 42 hours of psychotherapy. The GSR is interpreted as a measure of the anxiety of the patient, or his "mobilization" against any cue threatening punishment by the therapist, such as any cue which the patient has learned to perceive as evidence of low esteem. This anxiety presumably motivates varying subtle forms of resistance and defense against the feared punishment. Several alternative interpretations of the relation between permissiveness and GSR are rejected by various findings.

Two different measures of resistance are negatively correlated with the permissiveness of the therapist. However, the attempt to show that this relationship is mediated by the "anxiety" or "readiness" which the GSR measures is obscured by the finding that the GSR is related to the emotional significance of the

patient's speech, as well as to the permissiveness of the therapist.

#### REFERENCES

1. DARROW, C. W. The galvanic skin reflex (sweating) and blood-pressure as preparatory and facilitative functions. *Psychol. Bull.*, 1936, **33**, 73-94.
2. DITTES, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, **54**, 187-191.
3. DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100-107.
4. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
5. MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1-15.
6. ROGERS, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
7. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

Received January 4, 1957.

# THE INFLUENCE OF THE MOTHER'S PRESENCE ON CHILDREN'S DOLL PLAY AGGRESSION<sup>1</sup>

HARRY LEVIN AND VALERIE F. TURGEON  
*Cornell University*

**D**URING the course of socialization, the mother teaches her child skills, beliefs, values, and brings under control the child's free expression of his impulses. Among the impulses to be controlled, aggression presents special problems. For one, the very nature of training creates frustrations that are likely to result in aggressive behavior. Such acts of aggression must in turn be controlled, a process likely to instigate more of the change-worthy behaviors, a sort of vicious circle. The socializer thereby becomes an instigator to the very action she is trying to change.

In most instances, however, mothers do succeed in bringing under control the aggressiveness they deplore. Through the punishment of aggression, the conditions that instigate aggression also instigate the anticipation of the pain that has, in the past, accompanied such actions. We expect that when the inhibitions are greater in strength than the urges to be aggressive, the child will not behave directly in the disapproved manner. The mother, then, comes to both instigate and to inhibit the child's aggressions.

Assume that the training at home is the point of origin on a continuum of stimulus generalization. Following Miller's displacement model (4), the greater the difference of a given stimulus situation from the relevant aspects of the home, the greater the probability of the overt occurrence of aggression, within the limit that aggression is still instigated. Sears and his co-workers (e.g., 2, 7, 8) have conceptualized doll play aggression in these terms. The doll play situation is assumed to be similar enough to the home to instigate aggression yet different enough to escape some of the inhibitions.

The doll play setting has certain superficial similarities to the home: a doll family, house, and furniture. The differences are also important and designed to relax inhibitions against

aggression; the people are, after all, dolls, and the furniture and rooms are miniatures. Also, the experimenter is deliberately friendly and nonevaluative. The prediction has been that with inhibitions against aggression relaxed in doll play, the amount of aggression the child shows is a function of the degree of instigation, usually measured by how severely the child has been punished at home.

The present study asks how aggression in doll play is affected if the doll play setting is in some ways made more similar to the home by having the child's mother present at a doll play session. The expectation from the foregoing line of argument is that aggression should be reduced.

## METHOD

Since doll play usually involves a single adult interacting with the child, a control was necessary to insure that the child's doll play behavior would be attributable to the presence of the mother and not to the circumstance that two adults were present. Consequently, the control group's doll play was observed by an adult female, previously unknown to the child. The results from this study may also be compared with the several extant sets of norms of doll play behavior in which the preschool child participated only with the adult experimenter (3, 6).

The design, then, involved two groups of children, an experimental and a control group. For each group there was a control and an experimental session. In the first session, both groups were treated alike; only the experimenter was present. In the second session, the experimental group were observed by their mothers in addition to the experimenter, and the control group by a stranger and the experimenter.

## Subjects

The Ss were 20 children, 10 boys and 10 girls, aged three to five years. They came from upper middle-class, professional homes; 13 of the 20 were enrolled in the Cornell University Preschool. The children were first separated into groups of the same sex and age.

<sup>1</sup> This study was supported in part by a grant from the National Institute of Mental Health (M-901), "Pride and Shame in Children." We wish to thank Dr. Alfred L. Baldwin for his help with all aspects of the study and Drs. Urie Bronfenbrenner and John Harding for their careful readings of the manuscript.



Within each age-sex group the children were assigned randomly to the two treatments.

### Procedure

The Ss were observed in two sessions of standard, experimental doll play, each lasting 20 minutes. There was usually one day intervening between sessions, and in no case was the interval longer than three days. The doll play procedure has been described in detail elsewhere (6). A standard family of dolls was used: mother, father, boy, girl, and baby. The doll house had six uncovered rooms; the furniture was movable, but the walls were fixed. The same adult female (VT) was the experimenter in all sessions. The first session involved only the E and the child.

At the second session, the stranger-present (control) Ss entered the experimental room with the E to find an unfamiliar female adult in the room. This person acted as an observer who did not speak either to the E or the S, but simply sat watching the session.

The mother-present Ss entered the second session accompanied by their mothers and the E. The mothers had been previously instructed to watch the session but not to start interaction with the child. If the child talked to them they were to reply in a natural manner.

The sessions were observed from behind a one-way vision screen. The doll play record was unitized at the time of the observation into subject-predicate units which were recorded verbatim. "The boy runs," "the daddy spans the girl" are examples of single units. A scoring convention somewhat different from other doll play studies was adopted in that repetitive sequences such as the "boy runs and runs and runs" was scored as one unit. From the protocols two scores were derived: (a) total number of units per session and (b) per cent aggression per session, computed as the number of aggressive units divided by the total number of units per session.

Our attempt to score the agents and objects of the aggressive acts was not profitable. Usually, the child failed to specify either the agent or the object of the act.

Aggression was defined as any thematic action that irritates, hurts, injures, punishes, frustrates, or destroys a doll or equipment. It ranged in intensity from innocuous mischief and the violation of usual household rules (the boy jumps on the bureau) to severe accidents, fights, and catastrophes (the father breaks his leg, the boy has his head sliced off, a storm breaks up the house).

Verbal aggression such as scolding and threats were included, as were instances in which discomfort was attributed to a doll (the boy is sick, sad, lost, etc.).

Four sessions were observed by two observers. The four protocols were combined and agreement computed according to the formula: % Agreement = number of agreements/number of agreements + number of disagreements. For total units, the agreement between observers was 90%; for aggression units, the agreement was 83%.

### RESULTS

The score computed for the aggression analysis is the percentage of total units which are aggressive. This score controls for the overall level of general verbal thematic productivity of the child. For the various analyses, the percentage scores were normalized by use of the angular transformation,  $p = \sin^2 \phi$ .

To ascertain the comparability of the experimental and control groups, an analysis of variance of the aggression scores in the first session, prior to the experimental treatment, was performed. The difference in the average per cent aggression between the two treatment groups was not significant.

Since the results of the present experiment are to be compared with existing doll play norms, it is relevant to inquire whether the doll play behavior of the present Ss in the initial session diverges significantly from that of the groups on which norms are available. From a study of 202 boys and 177 girls, Levin and Sears (3, p. 149) report that the average per cent aggression scores in the initial doll play session is 14.28 for boys, and 8.64 for girls. The mean and .05 fiducial limits for the ten boys in this study is  $23.10 \pm 14.14$ ; and for the girls,  $6.90 \pm 4.98$ . While the obtained means are not as similar as might be desired, these findings are compatible with the view that doll play performance in this situation is typical of other, larger samples of children.

TABLE 1  
AGGRESSION SCORES IN TWO DOLL PLAY SESSIONS  
(N = 5 in each cell)

Sex	Mother Present <sup>1</sup>				Stranger Present			
	Session I		Session II		Session I		Session II	
	Mean*	SD	Mean	SD	Mean	SD	Mean	SD
Boys	16.40 ± 21.87	17.62	41.20 ± 36.48	29.39	29.80 ± 24.48	21.34	23.60 ± 20.71	16.69
Girls	6.40 ± 13.04	10.49	10.60 ± 17.49	14.08	7.40 ± 2.72	2.19	3.80 ± 7.25	5.85

\* The mean is followed in each case by fiducial limits at the .05 level.

TABLE 2  
SUMMARY OF ANALYSIS OF VARIANCE OF AGGRESSION  
SCORES IN TWO DOLL PLAY SESSIONS

Source	df	Mean Squares	F	p
Treatment	1	11.02	—	—
Sex	1	2666.68	9.30	<.01
Sessions	1	126.02	—	—
Treatment $\times$ Sex	1	1.77	—	—
Treatment $\times$ Sessions	1	746.50	7.66	<.02
Sex $\times$ Sessions	1	191.85	—	—
Treatment $\times$ Sex $\times$ Sessions	1	34.22	—	—
Error (b)	(16)	286.85		
Error (w)	(16)	97.41		

The questions under investigation concerning the influence of the mother's presence in the second session were approached through an analysis of variance of normalized per cent aggression scores according to the following classifications: treatment, sex of *S*, and session. Treatment, sex, and the interaction between the two involve the sums of the aggression scores for the two sessions. For these three sources of variation, the appropriate error term is the variance estimate, among *Ss*, within sex and treatment (labelled "Error b" in Table 2). The differences between the scores in the two sessions enter into the main effect of "sessions" and the various interactions between sessions and the other variables. Since the aggression scores in the two sessions are correlated, the error term for testing the differences between sessions is the interaction, *Ss*  $\times$  sessions, within sex and treatment (labelled "Error w" in Table 2).

For the aggression scores in both sessions combined, boys are more aggressive than girls. This finding replicates a number of other studies that reported sex differences in doll play behavior (1, 3, 5, 6).

The important results for the present study are the changes from Session 1 to Session 2. As can be seen in Table 2, the session  $\times$  treatment interaction is significant at about the .02 confidence level. Aggressiveness clearly increases from the first to the second session when the mother is watching the child and decreases when a stranger is present. Every one of the ten experimental children increased in aggressiveness when the mother was present. By a sign test, the probability of there being no reversals in a series of ten observations is .01.

Among the children in the presence of a stranger, the aggression scores of eight decrease in the second session; two children, one boy and one girl behaved differently from the majority.

The results in the experimental and control groups may be analyzed separately. Tested against a hypothesis of no difference, the increase of aggression in the group with the mother present is statistically significant ( $t = 3.34$ ,  $p < .01$ ). For the control group, the decrease is not significant ( $t = 1.00$ ,  $p < .30$ ).

The usual occurrence in doll play with only the *E* present is that aggression increases in the second over the first session. Is the increase in the present experimental group different from what might be expected in a normal second session without mother or stranger observing? According to available norms (3, p. 149) the mean increase in per cent aggression between sessions is 7.56 for boys and 4.12 for girls. For the present *Ss*, the differences between sessions and the .05 fiducial limits of these differences are  $24.80 \pm 30.37$  and  $4.20 \pm 4.91$  for the boys and girls in the experimental group and  $-6.20 \pm 28.04$  and  $-3.60 \pm 5.18$  for the boys and girls in the control group. The only group that differs from the scores expected on the basis of previous results is that of girls observed by a stranger, who appear to be inhibited in their aggressiveness.

## DISCUSSION

The prediction derived from displacement theory that the mother's presence would result in a decrease in her child's fantasy aggression was opposite to the results actually obtained. In light of the present findings, the displacement model appears to be an oversimplification.

In the first place, the theory assumes that in the parental treatment of aggression any frustration accumulates a potential for the child to be aggressive and any punishment for aggression tends to inhibit this behavior. Instigation and inhibition with respect to aggression are viewed as nonspecific, leading under appropriate conditions to behavior that is a resultant of the instigation and inhibition. This formulation may be scrutinized both in terms of how well it fits the complexity of aggression training and as to whether doll play, as used in the present study, provides adequate conditions for manifesting the behavior.



Recent evidence having to do with young children's aggression indicates that parents treat aggression differently depending on the nature of the aggressive acts and to whom the aggressions are directed. For example, Sears, Maccoby, and Levin (9) found, in a survey of child rearing practices, that mothers tend generally to be intolerant of aggressions directed toward themselves, somewhat more tolerant of fighting and quarreling among siblings, and still more accepting of their children's fighting with playmates. These distinctions are often sharp enough for the child to discriminate among aggressions, and to learn that under some conditions aggression is not at all condoned, whereas in other situations fighting is not only permitted but even encouraged. It may be that even with the mother present, children had categorized the laboratory situation as one appropriate to express their impulses, especially since their mothers had not unqualifiedly condemned aggression. On the other hand, it is not unlikely that one of the distinctions taught children vis-à-vis aggression is that the behavior is not appropriate in front of strangers. Such training might plausibly account for the decrease in aggressiveness in the stranger-present group.

One other discrimination that many parents probably reinforced could have influenced the behavior of the Ss in the presence of their mothers. It is quite unlikely that the children had been restrained from make-believe fighting or from having dolls and toys fight or be mischievous. Consequently, fantasy aggression may be considered by many children as approved regardless of whether or not their mothers are watching.

The above afterthoughts do not completely invalidate the displacement theory of fantasy aggression, but they argue for greater refinement than the theory now possesses. It would seem that the prediction of aggression in doll play could be improved if the fantasy situation were analyzed as to its similarity to approved or disapproved conditions of aggression. As we stated earlier, it is unfortunate that in the present study the agents and objects of aggressive acts were not specified frequently enough to warrant their separate analysis.

Two other hypotheses may be advanced to explain the fact that the presence of the mother did not restrain the children's aggressiveness.

One involves the presumption that the likelihood of retaliation for aggressive actions is a powerful inhibitor of aggression. It is soon clear to the child that there is no retribution in doll play, since the mother is in an obviously immobilized position. We may speculate that had the mothers been able to discipline their children, the amount of overt aggression would have been reduced rather than increased.

A second conjecture about the increase of aggressiveness in the mother-present group arises from the nature of superego development in the child.<sup>2</sup> Self control is a worrisome burden for the child so that he is ready to transfer his newly acquired control to his mother when she is available. Anecdotally, we know that mothers often complain that their children are so well behaved only when they are not around. If we assume that returning impulse control to the mother is most likely for children with the strongest superegos, we may indirectly test this "transfer of superego" hypothesis with the present data.

There is some evidence that, at the preschool ages, girls have more strongly developed superegos than do boys (9),<sup>3</sup> so that we might expect girls to be more influenced by the mother's presence than are boys. The sex difference analysis gives little support to this hypothesis. As can be seen in Table 2, girls are no differently influenced by the experimental treatment than are the boys.

#### SUMMARY

Twenty children were observed in two sessions of doll play. The first session involved the child and an experimenter. During the second session, the child's mother was an audience for one group and an adult female, not previously known to the child, watched the sessions of the second group.

As is usual in doll play, boys are more aggressive than girls. Each of the ten children observed by their mothers were more aggressive in the second than in the first session.

<sup>2</sup> We are indebted to Dr. John W. M. Whiting for suggesting this interpretation of the findings.

<sup>3</sup> Questionnaires to the mothers of the children in this study asked the same questions as (9) about signs of superego development: confessions of wrongdoing, looking sheepish or guilty and admitting transgressions when asked. These data confirm the earlier results, indicating that girls show more signs of superego than do boys.

Eight of the ten children watched by a stranger decreased their aggressions in the second session. The findings are discussed in terms of modifications of the displacement theory of aggression.

## REFERENCES

1. BACH, G. R. Young children's play fantasies. *Psychol. Monogr.*, 1945, **59**, No. 2.
2. HOLLENBERG, ELEANOR, & SPERRY, MARGARET. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, **1**, 32-43.
3. LEVIN, H. & SEARS, R. R. Identification with parents as a determinant of doll play aggression. *Child Develop.*, 1956, **27**, 135-153.
4. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155-178.
5. PINTLER, MARGARET H., PHILLIPS, RUTH, & SEARS, R. R. Sex differences in the projective doll play of preschool children. *J. Psychol.*, 1946, **21**, 73-80.
6. SEARS, PAULINE S. Doll play aggression in normal young children: influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**, No. 6.
7. SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychol.*, 1951, **6**, 476-483.
8. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135-234.
9. SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston: Row Peterson, 1957.

Received December 26, 1956.



# VERBAL CONDITIONING AS A FUNCTION OF EXPERIMENTER CHARACTERISTICS

ARNOLD BINDER, DAVID McCONNELL<sup>1</sup>, AND NANCY A. SJOHOLM

*Indiana University*

A NUMBER of experiments have confirmed Greenspoon's (5, 6) findings that various stimuli can serve as reinforcements for classes of verbal responses. It has thus far been reasonably convincingly demonstrated that such categories as "plural nouns" (5, 6), "nonplural words" (5, 6), "animals" (2), "statements of attitude toward an academic issue" (8), and "first person pronouns" (4, 7, 11) form response classes, and that such experimenter responses as "Mmm-hmm" (2, 5, 6), "Good" (4, 7, 8, 11), and "expressions of agreement" (12) are effective reinforcements in leading to an increase in the frequency of subjects' responses falling in certain of these categories. It is especially significant that, with few exceptions, the subjects had been unable to verbalize the relationship existing between their own responses and the reactions of the experimenter (so-called "unconscious learning" or "learning without awareness").

Modern day interviewers of many differing schools of thought stress the importance of the interpersonal aspects of the interviewing process. According to this viewpoint, the interviewer (whether psychotherapeutic or otherwise) is an important variable to consider, since the responses of the interviewee are influenced by that particular interviewer's physical and behavioral characteristics. The changes in interviewee's expressions of opinion concomitant with alterations in the characteristics of the interviewer have been shown in the studies reported by Cantril (3) and by Robinson and Rohde (9). Sarason (10) has perhaps been most articulate in his advocacy that any complete analysis of a clinical or experimental situation involving a subject and an experimenter (or examiner or interviewer) must include consideration of the manner in which the social and personal characteristics of the latter influence the behavior of the former.

The purpose of the present experiment was to study the effect of experimenter differences upon performance in a verbal conditioning

situation. The question at issue was: Would two experimenters differing markedly in physical and social characteristics produce different rates of learning in randomly selected subjects in an experiment similar to those in which minimum experimenter response has produced changes in the verbal behavior of a subject without his awareness? This experiment can, of course, only be considered preliminary since positive results would indicate the need for further research aimed at finding the specific effects of varying experimenter variables upon dependent learning variables.

The class of subject responses reinforced for this experiment was the use of "mildly hostile" words in sentences made up by the subject, and the reinforcement selected was the response "Good." The hostile word class was chosen because of the interesting incidental, interview-relevant comparison that might thereby be facilitated.

## METHOD

### *Subjects*

Two groups of Ss were run for the study, both selected by means of routing slips voluntarily signed in a few introductory psychology courses at Indiana University. The two groups consisted of 29 and 30 Ss, respectively.

### *Experimenters*

The first group (29 Ss) was run by an attractive, soft-spoken, reserved, young lady who was 5'½" in height and 90 pounds in weight. The experimenter for the second group (30 Ss) was very masculine, 6'5" tall, 220 pounds in weight, and had many of the unrestrained personality characteristics which might be expected of a former marine captain. Both were students at Indiana University. Perhaps more important than their actual age difference of about 12 years was the difference in their age appearance; the young lady could have passed for a high school sophomore while the male experimenter was often mistaken for a faculty member. Henceforth, for the sake of convenience, we shall refer to the former experimenter as *FE* and the latter as *ME*.

### *Materials and Apparatus*

Table 1 shows the two sets of words used in this experiment, one set consisting of "mildly hostile" verbs and the other of "neutral" verbs, all in the past

<sup>1</sup> Now at Ohio State University.

TABLE 1  
VERBS PRESENTED TO SUBJECTS

Neutral	Hostile
scribbled	harmed
snored	raged
saluted	sneered
paved	squabbled
heard	vexed
lost	blamed
bounced	criticized
charted	bickered
assigned	aggravated
shrugged	punished
paraded	irritated
recited	resented
enclosed	hissed
reclined	quarreled
misplaced	annoyed
enlarged	jeered
compiled	pestered
subtracted	disliked
emerged	scolded
regulated	scowled

tense. The words were selected from larger lists generously supplied to the experimenters by Dr. Arnold Buss. Buss had six judges sort a large list of verbs into one of three categories: "neutral," "mildly hostile," and "intensely hostile." A word was accepted in a particular class only if five of the six judges agreed on its category. No "intensely hostile" words were used in the present study.

Seven sets of 20 pairs, each pair consisting of one hostile word and one neutral word, were obtained by a process of random matching. The matching of hostile words with neutral words for each set of pairs was done without replacement so that the 20 pairs in each set contained every word. Each of the pairs was typed in capital letters on a 3" x 5" white card. The pairs were placed about 1½" from the bottom of each card, both words on a single line, and roughly centered with a blank space between them. The right or left position of hostile and neutral words was randomized.

About ¾" above the paired verbs on each card, again in a single line of type, were placed the three pronouns "she," "he," and "they" in capital letters. The order of pronouns on each card was randomly determined, although the same three were used on all cards.

The presentation apparatus consisted simply of a front board (10½" in height by 18½" in width) and two side boards (10½" x 19") attached to the front board and extending back to the left and right of the experimenter. The purposes of these boards were to conceal the activities of the experimenter and to support a slotted holder for the 3" x 5" cards on the upper center portion of the front board. The apparatus was placed on a table of such height that even the shortest *S* could see the head of the shorter experimenter, *FE*.

### Procedure

Before starting the trials of each new *S*, the experimenter shuffled each of the seven stacks of cards which were before him, but concealed from *S* by the boards

described above. There was a sample card in the slotted holder as each *S* sat down in the experimental room, with the verbs "walked" and "ran" typed in the lower portion and the pronouns "he," "they," and "she" in the upper portion. The following instructions were then read:

Sit in front of this stand in such a manner that you can easily read the words on the card before you. I am going to put other cards before you one by one. Each of these will be like the one before you with five words on it, three words in the upper part of the card and two in the lower part. For each card, I would like you to make up a sentence using one of the three upper words as the first word in the sentence and one of the lower two words as the second word in the sentence. You may use any one of the three upper words and either one of the two lower words in your sentence.

I will place a new card before you each time you make up a sentence using the words of the preceding card. Now, it doesn't matter whether the sentence you make up is long or short, or even if it's simple or complex. It is important that you respond promptly with the first sentence that occurs to you. Let me stress that I would like you to respond with the sentence which occurs to you immediately after the exposure of the word card without further analysis or selection.

After reading the instructions and answering any questions by rereading the relevant portions of the instructions, the experimenter presented the first stack card which happened to come out on top as a result of the preliminary shuffling. In order to find operant response level, the experimenter provided no reinforcement for *S* responses to either this initial card or any of the subsequent cards in the first stack. After completing the first stack in this fashion, the experimenter presented the first card of the second stack without any delay between stacks or any indication to *S* that he had begun another stack. If the sentence made up by *S* in response to this card included the hostile verb rather than the neutral verb (regardless of the particular pronoun used), the experimenter responded, "Good." The experimenter said nothing if *S* used the neutral verb. The experimenter then proceeded through the remaining 19 cards in the second stack, reinforcing the use of hostile words with "Good" and not responding to the use of neutral words. The remaining stacks were presented in the same way without pause or change in pattern when proceeding from the last card of one stack to the first card of the next stack.

At the conclusion of the above procedure, each *S* was asked questions to determine whether or not he had insight into the relationship between his responses and the reactions of the experimenter. Great care was taken to insure that both experimenters carried out identical procedures and reinforcement patterns.

### RESULTS

Each set of 20 cards was considered a block for analysis purposes, giving a total of seven blocks of trials for each *S*. The data for each *S* consisted of the number of sentences involving



TABLE 2  
MEAN NUMBER OF HOSTILE WORDS USED  
PER BLOCK BY ALL SUBJECTS OF THE  
TWO EXPERIMENTS

Experi- menter	No. of Subjects	Block						
		1	2	3	4	5	6	7
FE	29	8.8	9.1	10.3	10.7	11.2	11.4	11.6
ME	30	9.4	9.8	10.1	10.6	10.9	11.1	11.2

the use of hostile words given in each of the seven blocks. Table 2 contains the mean number of hostile words used in each block for the Ss run by FE and ME, respectively.

Before comparing the performance of the two sets of Ss, it was desirable to establish whether or not any learning occurred for the two groups taken separately. For this determination, one of the tests for trend recommended by Alexander (1), was used. The test used assumes a block-to-block fluctuation pattern common to all Ss but requires no assumption as to the equivalence of individual slopes. Table 3 summarizes the procedures applied to the data of each experimenter to test the hypothesis of no consistent trial-to-trial fluctuation or, in other words, of plateau performance throughout.

The *F*s derived from group slope are significant beyond the .001 level and indicate that the two groups tend to increase in a linear fashion. The insignificant *F*s derived from group deviations from linearity indicate that neither group shows consistent fluctuation apart from the linear trend.

But the question immediately arises: How many of the Ss could verbalize the relationship

between their behavior and the experimenter's responses, and would there be any consistent learning trend if only those Ss who showed no insight were included? Actually, seven of the Ss in FE's group and five of the Ss in ME's group demonstrated insight by verbalizing the contingency.

Table 4 contains the results of the same trend test as described above as applied only to those Ss who had no insight. There were, of course, 22 such Ss in the group of FE and 25 in ME's group. Again the *F* for group slope for the Ss of FE is significant beyond the .001 level. However, the corresponding *F* for the Ss of ME just misses being significant at the .05 level (an *F* of 3.92 is necessary). The *F*s derived from group deviations from linearity are once again not significant, indicating no consistent fluctuation other than linear trend. These results, together with the finding (reported below) of a highly significant over-all slope, support the hypothesis that the conditions of this experiment brought about the un verbalized learning to use hostile words rather than neutral ones in the structure of free sentences.

The primary hypothesis that the learning rates of Ss in verbal conditioning are a function of the personal characteristics of the experimenters, even though the participation of the experimenters is minimal and Ss are not aware of the experimenter response contingency, was tested by an extension of Alexander's test for trend. The extension tests the reasonableness of concluding that two groups differ in slope. Only those Ss who did not demonstrate insight in the questioning at the end of the experiment were used in this analysis. The results may be

TABLE 3  
SUMMARY OF TREND TESTS FOR THE TWO FULL GROUPS OF SUBJECTS TAKEN SEPARATELY

Source	Experimenter FE				Experimenter ME			
	SS	df	MS	<i>F</i>	SS	df	MS	<i>F</i>
Individual deviations from estimation (error)	807	140	5.76		519	145	3.58	
Group deviations from linearity	14.0	5	2.80	.49	3.00	5	.60	.17
Between individual slopes	298	28	10.64		384	29	13.24	
Group slope	200	1	200	34.72*	82.0	1	82.0	22.90*
Between individual means	2493	28	89.04		1214	29	41.86	
Total.	3812	202			2202	209		

\*  $p < .001$ .

TABLE 4  
SUMMARY OF SEPARATE TREND TESTS FOR THE TWO GROUPS OF SUBJECTS SHOWING NO INSIGHT

Source	Experimenter FE				Experimenter ME			
	SS	df	MS	F	SS	df	MS	F
Individual deviations from estimation (error)	548	105	5.22		427	120	3.56	
Group deviations from linearity	7.0	5	1.4	.27	1.0	5	.20	.06
Between individual slopes	144	21	6.86		151	24	6.29	
Group slope	113	1	113	21.65*	13.0	1	13.0	3.65
Between individual means	331	21	15.76		529	24	22.04	
Total	1143	153			1121	174		

\*  $p < .001$ .

seen in Table 5. The  $F$  coming from the between group slopes is significant beyond the .01 level, in support of the hypothesis that the two groups differ in learning trend. The significant  $F$  derived from between group means ( $p < .05$ ) would lead us to reject the hypothesis that the two groups do not differ in average number of hostile words. Fig. 1 shows the learning curves of these groups over the seven blocks.

The very highly significant over-all slope ( $p < .001$ ) requires rejection of the null hypothesis that Ss do not show unlearned learning to use hostile words under the conditions of this experiment.

The availability of Ss at the time of the experiment did not allow us to equate the two groups for numbers of males and females. Not only did FE start out with more males than ME, but five of the seven Ss eliminated (for

reasons of insight) from her group were females. Of the five eliminated from ME's group, three were male and two were female. Thus, for the test of difference in group slope, there were six females and 16 males in the FE's group, and 15 females and 10 males in ME's group. In order to justify any conclusions regarding the effect of experimenter characteristics upon learning, then, we must show that the differences found between the two groups could not be accounted for purely on the basis of the differences in sex composition.

Accordingly, the entire array of Ss without insight was divided into two groups on the basis of sex and irrespective of the particular experimenter used to get their data. Table 6 presents a summary of the Alexander test for differences in trend as applied to the comparison between female and male groups. The fact

TABLE 5  
SUMMARY OF THE TEST COMPARING THE TRENDS OF THE TWO GROUPS SHOWING NO INSIGHT

Source	SS	df	MS	F
Individual deviations from estimation (error)	977	225	4.34	
Between individual slopes	293	45	6.51	
Between individual means	860	45	19.11	
Group deviations from estimation	5.0	5	1.0	
Between group slopes	30.3	1	30.3	6.98**
Between group means	19.0	1	19.0	4.38*
Over-all deviations from linearity	5.0	5	1.0	.23
Over-all slope	93.6	1	93.6	21.57***
Total	2282.9	328		

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

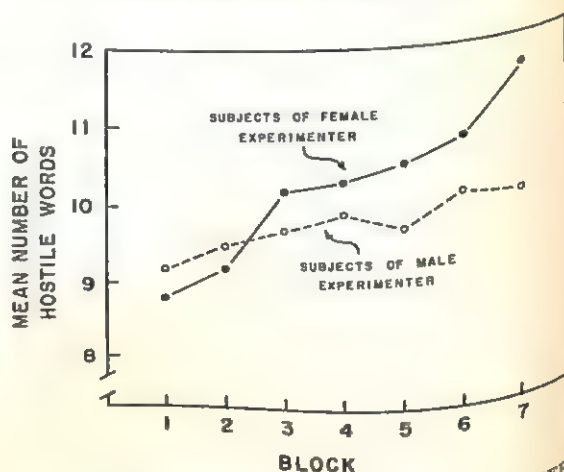


FIG. 1. LEARNING CURVES SHOWING MEAN NUMBER OF HOSTILE WORDS USED BY SUBJECTS WITHOUT INSIGHT IN THE TWO EXPERIMENTAL GROUPS



TABLE 6  
SUMMARY OF THE TEST COMPARING FEMALE  
AND MALE SUBJECTS WITHOUT INSIGHT

Source	SS	df	MS	F
Individual deviations from estimation (error)	958	225	4.26	
Between individual slopes	319	45	7.09	
Between individual means	877	45	19.49	
Group deviations from estimation	23.0	5	4.6	
Between group slopes	3.9	1	3.9	.92
Between group means	2.0	1	2.0	.47
Over-all deviations from linearity	5.0	5	1.0	.23
Over-all slope	93.6	1	93.6	21.97*
Total	2281.5	328		

\*  $p < .001$ .

TABLE 7  
MEAN NUMBER OF HOSTILE WORDS USED  
PER BLOCK BY MALE AND FEMALE  
SUBJECTS WITHOUT INSIGHT

Sex	No. of Subjects	Block						
		1	2	3	4	5	6	7
Male	26	8.7	9.1	9.6	10.5	10.0	10.6	10.5
Female	21	9.3	9.7	10.3	9.6	10.1	10.1	11.0

that neither the difference between group slopes nor the difference between group means led to significant  $F$ s supports the hypothesis that males and females do not differ in learning rate or over-all level of usage of hostile words in this particular context. The mean values for the two groups may be seen in Table 7.

#### DISCUSSION

The principal hypothesis of this study was that learning rate in a verbal conditioning situation is a function of experimenter characteristics even though the participation of the experimenter consists only of making a single word reinforcing response and the  $S$  is unaware of the reinforcing contingency. This hypothesis was supported by the experimental results. However, the direction of the difference in group slope is quite interesting. As may be seen in Fig. 1, the  $S$ s in  $FE$ 's group have a lower mean number of hostile words for the first (operant) block but a higher mean number for the last block as compared with the  $S$ s in  $ME$ 's group. It surely would not have been easy to predict that the  $S$ s of  $FE$  would show a significantly greater learning rate or slope than

the  $S$ s of  $ME$ . Preliminary conjecture, in fact, found it reasonable to expect that the response "Good" would have more of a reinforcing effect when uttered by the big, prestigious, male experimenter ( $ME$ ) than by the slightly built, soft-spoken, female experimenter ( $FE$ ). On an a posteriori basis, however, one may speculate that  $FE$  provided a less threatening environment, and the  $S$ s consequently were less inhibited in the tendency to increase their frequency of usage of hostile words.

The finding that the experimenter response of "Good" led to a significant trend of greater usage of hostile words in sentences is of some interest in itself, as it had not been demonstrated before. Buss and Durkee<sup>2</sup> carried out exploratory research on the verbal conditioning of hostile words. They presented their  $S$ s with cards containing a "neutral," a "mildly hostile" and an "intensely hostile" verb under instructions to make up a sentence using one of the three words on each newly presented card. These investigators found that little or no conditioning was achieved when the experimenter said "Good" after correct responses and nothing after incorrect responses, although conditioning did occur when "Right" and "Wrong" were used as reinforcers. Buss and Durkee, moreover, reinforced the intensely hostile words rather than the mildly hostile words in order to maximize the chances for learning.

There were certain differences in procedure between the Buss and Durkee experiment and the present one which might account for the differing findings on the reinforcing effect of "Good." But perhaps more important than these is the fact that different experimenters were involved in the two investigations—a factor which this experiment has shown to involve significant effects.

Some of the other differences reported in the literature on verbal conditioning may be accounted for on the basis of experimenter variability. Verplanck (12), for example, reported that some of his student experimenters obtained learning effects in their  $S$ s while other experimenters obtained essentially plateau performance.

While it is true that the differences in learning effects found in this experiment were

<sup>2</sup> Buss, Arnold H., & Durkee, Ann. Personal communication. November 21, 1956.

concomitants of markedly different physical and temperamental characteristics of the experimenters, such variability is not at all uncommon in psychological experimentation. Moreover, it would not be unreasonable to expect that greater differential effects would be obtained in experiments where the extent of experimenter participation is greater than in the present experiment. Future research could profitably be directed toward determining the specific personality and/or physical characteristics of experimenters which lead to differential learning effects, the types of learning situations in which such effects occur, and the constructs from learning and personality theory which can account for the phenomena.

#### SUMMARY

The purpose of this investigation was to test the hypothesis that differences in the characteristics of experimenters can lead to differential learning effects when the experimental session involves verbal conditioning without awareness. Two experimenters of different sex and markedly different height, weight, age appearance, and personality ran separate groups of Ss. The Ss were presented with 3" x 5" cards which contained the pronouns "he," "she," and "they" in random order and two past tense verbs, a "neutral" one and a "mildly hostile" one. The task of each S was to make up a sentence starting with one of the pronouns and utilizing one of the two verbs. A total of 20 neutral and 20 hostile verbs was used, arranged in seven blocks with random pairings of neutral and hostile words within the blocks. The experimenters said "Good" whenever an S used the hostile verb rather than the neutral one in his freely constructed sentence.

The female experimenter's group totaled 29 Ss, but seven of these were eliminated from the final analysis because they could verbalize the relationship between their behavior and the experimenter's response. The corresponding

numbers for the male experimenter's group were 30 and five.

The data were analyzed by means of trend tests recommended by Alexander (1). The results showed that the response "Good" was reinforcing for the class of behavior consisting of the use of hostile words in sentences. In addition, it was found that the rates of learning for the Ss of the two experimenters differed significantly, with a steeper slope for the female experimenter's group.

#### REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533-557.
2. BALL, R. S. Reinforcement conditioning of verbal behavior by verbal and non-verbal stimuli in a situation resembling a clinical interview. Unpublished doctoral dissertation, Indiana Univer., 1952.
3. CANTRIL, H. *Gauging public opinion*. Princeton: Princeton Univer. Press, 1944.
4. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106-110.
5. GREENSPOON, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctoral dissertation, Indiana Univer., 1951.
6. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409-416.
7. GROSSBERG, J. M. The effect of reinforcement schedule and response class on verbal conditioning. Unpublished doctoral dissertation, Indiana Univer., 1955.
8. HILDUM, D. C., & BROWN, R. W. Verbal reinforcement and interviewer bias. *J. abnorm. soc. Psychol.*, 1956, **53**, 108-111.
9. ROBINSON, D., & ROHDE, S. Two experiments with an anti-Semitism poll. *J. abnorm. soc. Psychol.*, 1946, **41**, 136-144.
10. SARASON, S. B. *The clinical interaction*. New York: Harper, 1954.
11. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496-501.
12. VERPLANCK, W. S. The control of the content of conversation: reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668-676.

Received January 18, 1957.



# MEDIATED GENERALIZATION AND THE INCUBATION EFFECT AS A FUNCTION OF MANIFEST ANXIETY<sup>1</sup>

MARTHA T. MEDNICK<sup>2</sup>

*Northwestern University*

ONE OF the earliest and most widely cited studies attempting to verify clinically observed phenomena in the experimental laboratory was reported by Diven (6). In order to test some notions about the nature of anxiety, he conditioned a psychogalvanic response (PGR) to the word "barn" and measured autonomic reactivity to other rural words. Responsivity to the rural words was found to be greater than to a group of nonrural words, and this was considered to be evidence for the generalization of anxiety. In addition, conditioned autonomic activity tended to increase as a direct function of the length of a rest pause preceding extinction. This was called the "incubation effect." White, Cameron, and others (20, 4, 11), in their discussion of anxiety, deal with these aspects of the Diven study at length.

The Diven findings on verbal generalization may be interpreted in terms of recent attempts to explain clinical phenomena on the basis of learning theory. The study may be regarded as an instance of mediated generalization (MG), a concept discussed by Cofer and Foley (5), which may be illustrated, using Diven's verbal stimuli, as follows:

Barn + shock.....	elevated PGR
Barn.....	elevated PGR
Cow (a rural word).....	Barn.....elevated PGR

"Barn," the implicit common or mediating term elicited by "cow," elicits an elevated PGR.

MG has been used in attempts to understand the role of covert symbolic processes in the determination of behavior. Dollard and Miller (7) placed considerable emphasis on this process in their discussion of personality de-

velopment in terms of learning theory. These writers also used the concept of anxiety incorporated into the Hullian system by assuming that an anxiety response has drive properties and contributes to the total drive state of an organism. Some verification of this hypothesis concerning the drive properties of anxiety has been achieved through studies of the performance on learning tasks of human Ss separated on the basis of response defined anxiety by means of extreme scores on the Taylor Manifest Anxiety Scale (MAS) (17). The results of a number of studies have supported predictions based on the assumption that MAS scores are related to drive level (18). Dollard and Miller argued further that generalization varies directly with drive and, hence, with anxiety. While stimulus generalization (generalization along a physical continuum of similarity) has been shown to vary with primary drive in animal studies (3, 12) and with anxiety (13), such a relationship has not been clearly demonstrated with MG. Lacey (9), replicating Diven's study and results, found that generalization tended to vary directly with MAS scores, but the group differences did not reach significance. The Lacey study must be interpreted with caution since the conditioning procedures utilized were somewhat irregular; the CS-UCS interval was 15 sec., filled with S associating aloud to the CS. This method was utilized in the interest of maintaining S's unawareness of the CS-UCS contingencies.

The incubation effect described above was the other major finding reported by Diven. Bindra and Cameron (1) also studied this effect during the presentation of noxious stimulation rather than under conditioning and extinction paradigm. These investigators concluded that there appeared to be a genuine incubation effect but that little is known of its exact nature or its relation to other variables.

The present study was an attempt to reproduce Diven's and Lacey's findings with respect to the generalization and incubation of a conditioned autonomic response and to examine

<sup>1</sup> This paper is based upon a dissertation submitted to the Department of Psychology of Northwestern University in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to thank Prof. Janet A. Taylor, dissertation sponsor, for her advice and criticism. Mr. Bernard Bernstein gave much needed technical assistance during the pilot stages of this study. This paper was written while author was a USPHS Post-doctoral Fellow, Harvard Medical School.

<sup>2</sup> Now at 63 Madison Rd., Waltham, Massachusetts.

the relationship of mediated generalization to manifest anxiety. An emphasis was placed on an empirical choice of generalization stimuli. It should be noted that generalization words were chosen by Diven and Lacey on a rational basis. It was felt that additional work in this area would profit by a pre-experimental demonstration of the associational relationship of the generalization stimuli to the CS by making it possible to order the stimuli along a rough continuum and to attempt to obtain a gradient of MG.

## METHOD

### Subjects

Two groups of 45 Ss each were chosen from introductory psychology classes at Northwestern University on the basis of extreme scores on the Heineman form of the MAS (8). Forty-five individuals (15 males and 30 females) receiving scores of 64 and above on the scale were included in the high anxious group (HA), and the 45 Ss scoring 49 or below (35 males and 10 females) were designated as the low anxious group (LA). In addition, three LA and four HA Ss were tested, but because they showed no evidence of conditioning, they were excluded from further consideration. Ss were assigned to the various conditions by a noninvolved person so that E had no knowledge of MAS scores during the experiment.

### Apparatus

Galvanic measures were made from a standard Stoelting polygraph. Basal readings, taken directly from a dial on the instrument, were recorded periodically. At random intervals during the presentation of the experimental stimuli, an artificial resistance of 1000 ohms was thrown into the machine. This provided a standard measuring unit for each record. The verbal stimulus materials were presented to S by means of a tape recorder via earphones. A loud raucous noise which served as the UCS was also presented by means of the tape recorder. A measure of the intensity of the test stimuli revealed no differences between stimuli.

### Verbal Stimulus Materials

In constructing the stimulus materials, the results of a recent study by Russell and Jenkins (15) were utilized. These investigators administered the 100 words from the Kent-Rosanoff Word Association Test to 1,007 college students. A single association was elicited to each of the words and the frequency of each association determined. For the present study, the word that was chosen as the CS appeared (in varying degrees of frequency of association) as a response to a number of different Kent-Rosanoff stimulus words. Thus, the words "dark," "heavy," "soft," and "lamp" all had as a common response the word "light." "Light" was therefore designated as the CS in the present study and the words to which it had been associated as the generalization test words.

TABLE 1  
PERCENTAGE OF INDIVIDUALS ASSOCIATING  
THE WORD "LIGHT" TO GENERALIZATION  
WORDS

Sample	N	Generalization Words			
		Dark	Lamp	Heavy	Soft
Minnesota	1007	83	63	58	.86
Northwestern	174	66	59	47	.1
Minnesota <sup>a</sup>					
HA	28	86	50	54	0
LA	28	82	61	57	0
Northwestern					
HA	25	60	60	32	4
LA	25	64	60	60	0

<sup>a</sup> Personal communication from Dr. Shirley Johnson, University of Minnesota.

Table 1 presents the percentage of individuals in the Minnesota group who responded to each generalization test word with the word "light." A list of 40 of the 100 Kent-Rosanoff stimulus words containing all of the test words was also administered to a sample of Northwestern University students; the percentage of individuals responding with the word "light" to the test words in this group can also be seen in Table 1. Since it is important to show that HA and LA individuals do not differ from each other with respect to these associational frequencies, the relevant data are included in Table 1. No important differences were found between these two groups.

The stimulus materials used during the conditioning trials consisted of the CS and a list of 24 neutral words (a neutral word being one to which the CS was not given as an association by the Minnesota group). Each of the neutral words was repeated three times in differing orders, while the critical word, "light," was presented on 14 occasions. Thus, the training series consisted of 72 presentations of neutral words and 14 randomly spaced presentations of the CS. The raucous buzzer followed the CS on 9 of the 14 occasions it was presented. This partial reinforcement procedure was followed for two reasons: as an attempt to prolong extinction, and to allow the trials on which the CS was presented without the UCS to serve as a test for conditioning. The test stimulus list, used for measures of generalization and extinction, consisted of the CS, the four generalization stimuli, and three neutral stimuli. This series was repeated eight times with the order of the four generalization words and one of the neutral words ("square") being counterbalanced so as to avoid positional effects. The remaining two neutral words were included so as to increase the number of words heard by S between the other words. The time between the presentation of each word in both the conditioning and test series varied from 8 to 12 sec., thus allowing some time for return of S's response to a basal level.

### Conditioning Trials

The Ss in all groups received the same conditioning procedure. The experiment was conducted in a semi-soundproof room that offered relative freedom from



distraction. The earphones effectively eliminated other distracting noises. Each *S* was seated in a comfortable broad-armed easy chair with electrodes and earphones attached to him. Initial instructions, designed to acquaint *S* with the nature of the apparatus, were given by *E*. A cloth screen shielded the apparatus from *S*'s view. Additional instructions informing *S* about the stimuli to be presented were on tape and heard by *S* over the earphones. These indicated that a buzzer and words were to be presented at various intervals and that *S*'s task was merely to remain quiet and listen since his physiological reactions were being studied. No mention was made of the relationship of the buzzer to the appearance of any specific word.

Following the instructions, 15 "adaptation" words were presented to allow *S* to adjust to the apparatus before conditioning was attempted. These adaptation words were immediately followed by the conditioning series.

### Test-Extinction Trials

Each of the two anxiety groups was divided at random into three subgroups for the extinction procedure. One subgroup (Condition I) underwent a regular extinction procedure immediately following the conditioning series. There was no interruption between training and extinction, no additional instructions were given, and *S*s remained attached to the apparatus. The second subgroup (Condition II) had the electrodes and the earphones removed with instructions from *E* that they would be allowed to rest; *S* remained in the chair with *E* seated out of sight behind the screen. After a lapse of 10 min., *S* was reattached to the apparatus for the extinction procedure. The *S*s were readapted to the point where deflections had mainly ceased before the tape recording was again turned on. The third subgroup of *S*s (Condition III) was allowed to leave following training and instructed to return exactly 24 hrs. later for the remainder of the experiment. Prior to his final dismissal, each *S* was asked to report whether he knew which word had been followed by the buzzer.

### RESULTS

Raw ohms resistance data were obtained from the records by measuring the heights of galvanic deflections, using the 1000 ohm artificial deflection as the measuring unit. Following the method suggested by Lacey and Segal (10), ohms resistance was then transformed into conductance units, micromhos (*M*), by use of the formula,  $M = 10^6 (1/r)$ . Due to the skewed nature of the data, median measures and appropriate nonparametric statistics were used throughout the analyses. In cases where analysis of variance by ranks was used, the obtained chi-square values were converted to *F* values (19).

As noted above, there were unequal proportions of males and females in the two anxiety

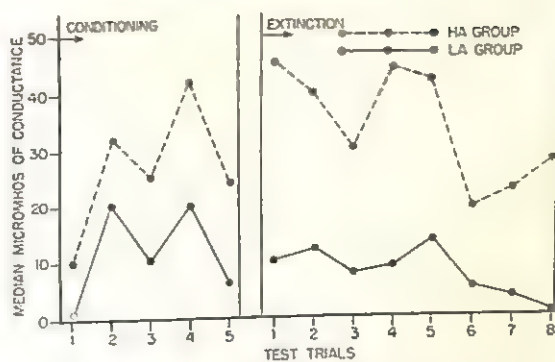


FIG. 1. MEDIAN AMPLITUDE OF CONDITIONED RESPONSES AT EACH TEST TRIAL DURING CONDITIONING AND DURING IMMEDIATE EXTINCTION (EXTINCTION DATA FOR CONDITION I ONLY)

groups so that the anxiety and sex variables were confounded. In order to evaluate the possible effects of sex on conditioning, subgroups of 23 males and 23 females equated for MAS scores were compared as to the amount of conditioning demonstrated. The differences were not significant (chi square = 1.8). There was thus no evidence that sex was a variable influencing conditioning in this study.

Another preliminary analysis was made in order to see whether the three incubation groups were equally conditioned before the commencement of extinction. An analysis of variance by ranks demonstrated that these groups did not differ significantly ( $F = .848$ ).

It should also be noted that of the 90 *S*s, 89 were able to state that "light" had been the word followed by the buzzer.

### Conditioning

The responsivity to the five conditioning test trials (those occasions on which the CS was not accompanied by the buzzer) represented a measure of conditioning. As can be seen from Fig. 1, the HA group responded with greater amplitude on all test trials than did the LA group. A total conditioning score for each *S* was obtained by summing the amplitudes of all five of his conditioned responses (CRs) during training. The differences between groups in total scores, evaluated by means of chi square, were significant (chi square = 4.8,  $p < .05$  for 1 *df*).

### Generalization

The tests for generalization were based on the amplitude data of the first extinction series

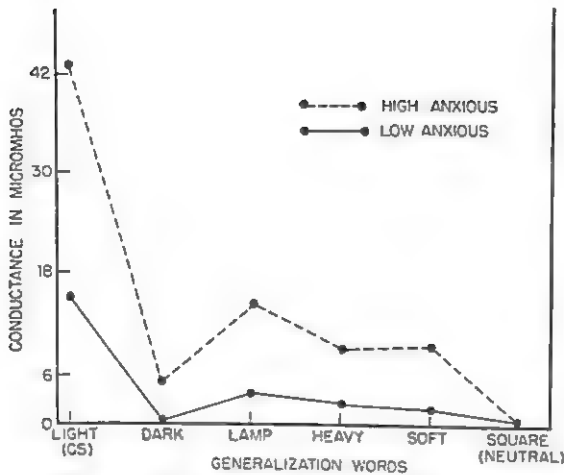


FIG. 2. MEDIAN AMPLITUDE OF RESPONSE TO EACH WORD FOR HA AND LA Ss

TABLE 2  
CHI-SQUARE ANALYSES BETWEEN WORDS  
FOR ALL Ss COMBINED  
(All at 1 *df*)

Word Pair	$\chi^2$	<i>p</i>
Light-Lamp	17.28	< .001
Lamp-Dark	5.28	< .05
Lamp-Heavy	2.78	n.s.
Lamp-Soft	2.37	n.s.
Dark-Square	4.65	< .05
Heavy-Square	10.32	< .01
Soft-Square	12.00	< .01

because of the great increase in the number of response failures in the later series. The word "square" was used as the neutral stimulus in all comparisons since the order of presentation of this neutral word was counterbalanced. As may be seen in Fig. 2, the median amplitude for the test words tended to be higher than for the neutral word in both groups. With respect to the degree to which the test words were associated with the CS "light," the most strongly associated word "dark" elicited less responsivity than would be predicted on the basis of association alone. The remaining three stimuli roughly followed the expected pattern, with "lamp" having the greatest reactivity.

In order to test the significance of the observed differences in responsiveness between the individual words, the data for all 90 Ss were combined and the differences in reactivity to the test words evaluated by means of chi square. The resulting values are reported in Table 2. All the associated words elicited significantly greater reactivity than did the

neutral word "square," thus demonstrating that generalization had taken place.

Turning to the relative performance of the two groups, examination of Fig. 2 reveals that the HA group showed consistently greater median amplitudes than the LA group. In order to obtain a total generalization score the amplitudes of responses to all four test words were summed for each S. A chi square with respect to this overall score demonstrated the HA group to have significantly more generalization reactivity than the LA group (chi square = 9.3,  $p < .05$  for 3 *df*).

Since the PGR amplitude of the HA Ss was significantly greater than the LA group during conditioning, this fact might be invoked to explain the obtained differences in generalization. In order to see whether these differences in generalization between the HA and LA groups could be demonstrated independently of this initial difference in conditioning level, two subgroups of 24 Ss each, one from the HA and one from the LA group, were matched on the basis of their total amplitude score in conditioning. A chi-square analysis between these subgroups of total generalization scores proved to be significant (chi square = 14.96,  $p < .01$  for 2 *df*).

A final comparison was made among the incubation groups. These three groups, it will be recalled, all received identical training and differed only in the amount of time allowed to lapse before commencing with the regular test-extinction procedure. As previously mentioned, it was demonstrated that these groups, referred to as Condition I (no rest interval), Condition II (10-min. rest interval), and Condition III (24-hr. interval) were equally responsive at the end of training; any subsequent differences are thus attributable to the rest-interval variable. Fig. 3 summarizes the generalization data of these three conditions for the HA and LA groups combined. An analysis of variance by ranks, testing the differences between Conditions I, II, and III, was significant ( $F = 15.0$ ,  $p < .01$  for 2 and 8 *df*). Inspection of Fig. 3 shows that this significant difference may be attributed to the consistently greater amplitude of Condition III, Conditions I and II being more or less similar at each of the generalization test points. It seemed possible that the amount of incubation time might affect the HA Ss differently from the LA Ss. The same



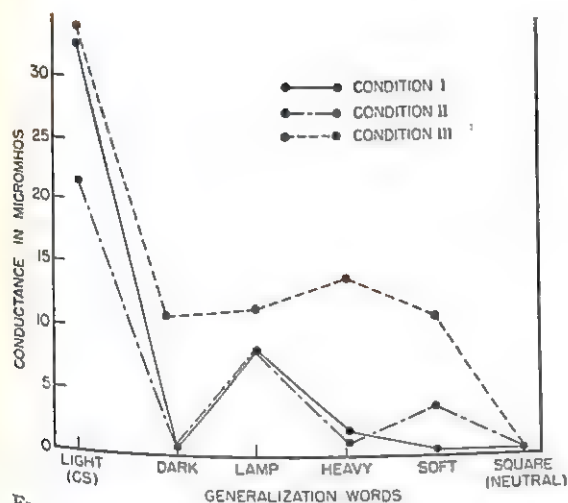


FIG. 3. MEDIAN AMPLITUDE OF RESPONSE TO EACH WORD FOR EACH CONDITION, HA AND LA Ss COMBINED

test between incubation conditions was therefore repeated for the LA and HA groups separately. For the HA group, the difference between Conditions I, II, and III did not reach significance ( $F = .455$ ), but significance was reached for LA Ss ( $F = 8.5$ ,  $p < .05$  for 2 and 8  $df$ ). The significance of differences between conditions reported above could thus be traced in large part to differences in the LA group.

### Extinction

The generalization data reported above represented the first of the eight series of test trials. The nature of all eight series makes it possible to follow the extinction of the CR. Only extinction of responses to the CS will be described since the test words extinguished very rapidly following the first series. Table 3 presents the median PGR amplitude at each extinction series for the CS for all groups and subgroups in the study. When the total HA and LA groups are considered, it may be seen that the HA group tended to maintain greater responsivity than the LA group until both groups finally demonstrated a median responsivity of zero. This greater responsivity was found to some degree in all three extinction conditions, but was most marked in Condition I where extinction immediately followed training. It might also be noted that contrary to the other conditions, the median PGR amplitude of the Condition I HA Ss did not approach zero by the end of the series, indicating that the CR did not completely extinguish.

TABLE 3  
MEDIAN AMPLITUDE AT EACH EXTINCTION SERIES FOR INCUBATION AND ANXIETY SUBGROUPS

	Extinction Series							
	1	2	3	4	5	6	7	8
Cond. I								
LA	10.0	12.0	7.5	13.0	13.5	5.0	3.5	1.0
HA	45.0	40.0	30.0	44.0	42.5	19.5	23.0	28.0
Total	32.5	22.0	13.0	28.0	23.0	6.0	13.0	10.0
Cond. II								
LA	19.0	11.5	23.0	12.0	7.0	2.5	3.5	0
HA	50.0	21.0	9.5	0	9.5	0	0	0
Total	21.5	13.0	19.0	8.0	8.0	0	0	0
Cond. III								
LA	37.0	17.0	8.0	9.0	3.5	0	0	0
HA	29.0	25.5	11.5	8.0	2.5	0	0	0
Total	34.0	22.0	9.0	9.0	4.0	0	0	0
Total Group								
LA	15.0	12.5	9.0	11.0	8.0	4.5	2.0	0
HA	43.0	33.5	20.5	17.5	10.5	1.0	0	0

### DISCUSSION

The results of this study confirm Diven's and Lacey's findings with respect to generalization. An inspection of the generalization data also seem to indicate that reactivity followed a rough gradient, i.e., autonomic responsivity varied directly with the degree of association of the test stimuli to the CS ("light").

Turning to the conditioning data, some comment may be made regarding the conditioning curves themselves. Although it is apparent that conditioning took place, the curves (Fig. 1) are irregular, responsivity being less at the end than in the middle of training. Such generally declining response curves have also been reported by Bitterman and Holtzman (2) and may be at least partly a function of adaptation. In the present study, an explanation in terms of the training procedure also suggested itself. It will be recalled that the CS was presented alone without the buzzer on irregularly spaced trials during training. Thus, conditioning test Trial 1 followed one UCS; two presentations of the UCS intervened between Trial 1 and Trial 2, one UCS between Trials 2 and 3, three between Trials 3 and 4, and one between Trials 4 and 5. Trials 2 and 4, consequently, had two and three UCS presentations immediately preceding them, whereas Trials 3 and 5 had only one. Inspection of Fig. 1 reveals that the irregularity is contributed to by a decline in responsivity on Trials 3 and 5. It is possible that the decline

of responsivity on these trials was at least partly related to their having fewer intervening presentations of the UCS; that is, presentation of the CS alone might have acted as an extinction trial, the effects of extinction still being noticeable when only one presentation of the UCS intervened between two test trials.

With respect to the anxiety variable, previous research using the MAS has shown that HA Ss exhibit both a higher level of performance in conditioning and extinction and higher gradients of SG (2, 18, 13). To the extent that SG and MG are affected by the same variables, HA Ss were expected to show more MG than LA Ss. The expected conditioning and extinction differences were found, and the expectation with respect to MG was also confirmed. It is particularly of interest, in view of the groups' differences in conditioning, that generalization differences were sustained between HA and LA subgroups equated for conditioning performance level. This supports the Dollard and Miller treatment of MG and SG as subsidiaries of the class of phenomena called generalization.

The extinction data of the group as a whole are somewhat at variance with those of other investigators (2) who reported consistently greater PGR responsivity during extinction for HA Ss than LA Ss. In the present case, this difference held up only through the fifth extinction series, after which it diminished sharply. However, it is important to note that Group I was the only group receiving immediate extinction and hence is the only one directly comparable to those of the earlier studies. An examination of Fig. 1 reveals that under Condition I, HA Ss showed greater responsivity throughout and also more resistance to extinction than the LA Ss. This result, then, supports the earlier finding. The data also suggest that an unsuspected effect of the rest interval was to speed extinction for the HA group and thus diminish the differences between the two anxiety conditions.

The incubation of anxiety has been accepted as a fairly meaningful and reliable clinical phenomenon. If an individual learns an anxiety reaction with respect to a particular stimulus situation and after a long interval returns, as it were, to the scene of the provocation, the anxiety is thought of as becoming more and more enhanced with the passage of time. As indicated, there is some experimental evidence to

support the assertion of such an effect. However, the findings of this study are not completely in line with this contention. In keeping with others' results, a 24-hr. interval between training and extinction operated to increase galvanic responsivity. This effect was short-lived, however, lasting only for the first extinction series. The 10-min. interval did not result in any PGR increase. This suggests only that recovery from adaptation may be involved and that a 10-min. interval is an insufficient amount of time for the recovery to show its effect.

MG has been demonstrated to be useful in the experimental study of learning and thinking (16). The main finding of this study, the demonstration of a gradient of MG and its relationship to manifest anxiety, indicates that this may also be a useful concept in the study of pathological disturbances of thinking and learning.

#### SUMMARY

The present study was designed to investigate the phenomena of mediated generalization and the incubation effect. An investigation was also made of the relationship of mediated generalization to level of manifest anxiety as measured by the Taylor Manifest Anxiety Scale (MAS).

Two groups of 45 Ss each were chosen on the basis of extreme scores on the MAS. A conditioned PGR was first established by repeated pairings of a critical word (CS) with a raucous buzzer. Tests of mediated generalization were made following training, using as generalization stimuli, words to which the CS was associated as determined by free associations of a standardization group. In order to investigate the incubation effect the Ss were divided following training into three subgroups differing in the length of the rest interval between conclusion of training and the institution of extinction trials (immediate extinction, 10-min. interval, and 24-hr. interval).

The results indicated heightened responsivity towards words associated with the CS, thus indicating that mediated generalization had taken place. It was also found that extreme MAS scores were directly related to the magnitude of mediated generalization responsivity. In agreement with reports of other investigators, level of manifest anxiety was found to be directly related to conditioning



responsivity as well as to resistance to extinction.

Increased PGR reactivity was found following the 24-hr. pre-extinction interval (i.e., an incubation effect), but the influence of this effect was short-lived being confined to the early part of extinction. The 10-min. delay prior to extinction did not result in galvanic activity which was significantly greater than that found with the immediate extinction condition.

# REFERENCES

1. BINDRA, D., & CAMERON, L. Changes in experimentally produced anxiety with the passage of time: incubation effect. *J. exp. Psychol.*, 1953, **45**, 197-203.
2. BITTERMAN, M. E., & HOLTZMAN, W. N. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952 **47**, 615-623.
3. BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209-226.
4. CAMERON, N., & MAGARET, ANN. *Behavior pathology*. Cambridge, Mass.: Riverside Press, 1951.
5. COFER, N., & FOLEY, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513-540.
6. DIVEN, K. E. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291-308.
7. DOLLARD, J., & MILLER, W. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
8. HEINEMAN, C. E. A forced choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447-454.
9. LACEY, J. I., SMITH, R. L., & GREEN, B. A. Use of conditioned autonomic responses in the study of anxiety. *Psychosom. Med.*, 1955, **17**, 208-217.
10. LACEY, O. L., & SIEGAL, P. S. An analysis of the unit of measurement of the galvanic skin response. *J. exp. Psychol.*, 1949, **39**, 122-127.
11. MURRAY, H. A. *Explorations in personality*. New York: Oxford, 1938.
12. ROSENBAUM, G. Temporal gradients of response strength with two levels of motivation. *J. exp. Psychol.*, 1951, **41**, 261-267.
13. ROSENBAUM, G. Stimulus generalization as a function of clinical and experimentally induced anxiety. Unpublished doctoral dissertation. State Univer. of Iowa, 1950.
14. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally produced anxiety. *J. exp. Psychol.*, 1953, **45**, 35-43.
15. RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word and Association Test. The role of language in behavior. *ONR Tech. Rep.*, No. 11, 1954.
16. RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired associate learning. The role of language in behavior. *ONR Tech. Rep.*, No. 10, 1954.
17. TAYLOR, JANET A. A personality test of manifest anxiety. *J. abnorm. soc. Psychol.* 1953, **48**, 285-290.
18. TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.* 1956, **53**, 303-320.
19. WALKER, H. M., & LEV, J. *Statistical inference*. New York: Henry Holt, 1953.
20. WHITE, R. W. *The abnormal personality*. New York: Ronald Press, 1948.

Received January 30, 1957.

# A SCALE FOR MEASURING ATTITUDES OF SOCIAL RESPONSIBILITY IN CHILDREN<sup>1</sup>

DALE B. HARRIS

*Institute of Child Welfare, University of Minnesota*

EARLIER publications (3, 4) have described an attempt to evaluate an attitude of social responsibility in children. One of the scales used, adapted from the Gough responsibility scale (2), and identified as the "How I Feel About Citizenship" scale has shown virtually no increment with increasing age from 9 to 18 years. This fact, rather unique in the child psychology literature, together with the fact that a number of the items had no theoretically demonstrable relationship to the definition of social responsibility adopted for this study, suggested that the scale be revised and revalidated. For the new instrument, we retained items from the old scale and devised additional items that could be related to our theory of responsibility as a composite of attitude elements reflecting behavior classifiable as reliable, accountable, loyal, or doing an effective job. A scale of 89 items was finally prepared, each of which could be justified as representing one or more of the above mentioned attitudes. Each item was answered as "agree" or "disagree."

## DEVELOPMENT OF THE SCALE

To provide a criterion, a pupil nomination form of a modified Guess Who type was administered to children in 13 seventh- and eighth-grade classes. The form asked each pupil to nominate the boys and girls who were "best fits" in his class for each of the four aspects of responsibility mentioned above.

The following untitled four descriptions, designed to illustrate these four features of responsibility, were submitted to the children:

1. This is a person upon whom you can always count; you can depend on him. When he promises that he will do something, you

can always count on his doing it. You can count on his word and trust him.

2. This person is a square shooter. He doesn't try to take advantage of or cheat others. When he has done something wrong, he will own up to it rather than try to blame some one else.

3. This person thinks for the good of others rather than always for his own gain. He is loyal to the group.

4. This person is one who gets things done. On a class project, on a committee or on a work job he gets right to work and can be counted upon to do well and promptly.

Below each of the descriptions, spaces permitted writing in three boys' names and three girls' names, members of the class who were considered "best fits" for each description. The children were not asked to write their nominations in order of merit. Children were instructed that any one child might be nominated for more than one characteristic. Independently of the children, the room teachers also filled in a nomination form. The 89-item scale of attitudes, now called the "Social Attitudes Scale," was administered to the same school classes a day or so following the "pupil nomination" form.

The children, who had spent all their school time with their principal teachers for the previous six months, were as well acquainted with each other and with their teachers as is generally found in public schools. They represented the socioeconomic distribution of the city. Two hundred and twenty-seven boys and 228 girls completed all of the schedules. Fortunately there were few absentees or incomplete records, and none among the crucial cases selected by the nomination form.

Although children were not required to nominate three of each sex, most of them supplied six names for each characteristic. A tally sheet was prepared for each of the four characteristics within each class group. The number of times a given child was mentioned for one characteristic constituted his score on that trait. Thus, the score did not distinguish whether a child's name appeared in first or

<sup>1</sup> This investigation was supported in part by a research grant (M-690) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, John E. Anderson and Dale B. Harris, Principal Investigators, and in part from funds supplied by the Institute of Child Welfare, University of Minnesota. The writer wishes to acknowledge the assistance of Dr. John E. Anderson and Dr. Harry Beilin in evaluating the theoretical relevance of items for the scale.



third place on a particular blank. Only a very few children received no mentions at all in their groups and, as is usually the case with Guess Who devices, five or six children in a group received the majority of the votes cast.

Rank-order correlations among the several traits for six class groups, one from each of six schools, show that the same names tended to appear high on the list for each of the four traits. For particular pairs of traits the medians of coefficients calculated from six school samples varied from  $+0.78$  to  $+0.82$ , showing the order of relationships among the four traits to be much the same regardless of the pair of traits selected. Moreover, there was relatively little variation among the several school classes in the size of obtained correlations; values ranged typically from  $+0.70$  to  $+0.89$ . The intercorrelations of scores based on one trait with scores obtained by summing scores for the other three characteristics ranged from  $+0.78$  to  $+0.95$  in various classes of 33 to 37 children.<sup>2</sup> Thus, no one characteristic appeared to be more closely related than another to the sum of scores on the remaining three aspects of responsibility. A total score obtained by summing the four trait scores appeared justified by this marked commonality among the aspects of responsibility.

To provide a criterion for selecting the attitude items associated with responsibility, 25 boys and 25 girls scoring highest on the sum of the four characteristics were selected from the total population. In every case children were selected who were also named for some component of responsibility on the teacher's ballot. In only three instances it was not possible to include the highest scoring child by the peer ballot when the teacher's judgment was added. These 50 responsible children were selected, two to six from each class, so as to include about a 10 per cent sample of each class enrollment. Boys and girls were selected in proportion to their distribution in the class. In the largest class groups an extra boy or girl was taken, since somewhat more than a 10

per cent sample was needed to produce 50 cases.

From each class an equal number of boys and girls were selected from the unmentioned or lowest scoring children, to constitute a criterion sample of "less responsible" children.<sup>4</sup>

### DISCUSSION

The attitude tests for these criterion groups were segregated and analyzed. Proportions of each group subscribing to each item were read onto Lawshe's nomograph (6) and the discriminatory power of the item recorded. Fifty items satisfied a Lawshe value of  $+0.4$  or better and have been retained. This represents approximately the 10 per cent level of significance. Of the 50 items, 21 appeared in the "How I Feel About Citizenship" scale used in the previous studies, and 29 were new items. Twenty-two of the items in the previous blank were not used in this schedule, either because they did not in our opinion have construct validity or did not hold up in the revalidation procedure described in this paper. This new plank, entitled "Social Attitudes Scale," was used in the 1954 Nobles County studies.<sup>5</sup> Retest reliabilities after an interval of four months ranged from  $+0.60$  to  $+0.70$  for various groups of eighth-and tenth-grade children. The items are reproduced in Table 1.

In another study, Rosemary Conzemius (1) used the "Social Attitudes Scale" with all 10-year-olds and high school seniors enrolled in the rural schools of another Minnesota county. She divided the children into those who belonged to 4-H Clubs, and those who did not. Table 2 compares data from her study with our results. Her non-4-H members tend to be slightly above our Nobles County rural youth, and her 4-H Club members appreciably exceed the Nobles County youth in comparable age groups. Conzemius found that 4-H Club members significantly exceeded the nonmem-

<sup>4</sup> This method does not select *irresponsible* children by nomination for specifically described characteristics. Direct peer nominations for socially unacceptable behaviors sometimes have unfortunate social and public relations consequences. It is assumed that most such children would be included in the low count or unmentioned children when the positive nomination is requested.

<sup>5</sup> A county-wide survey of adjustment and mental health in children was conducted through funds supplied by the National Institute of Mental Health and the Institute of Child Welfare.

<sup>2</sup> Tables giving these data in full have been deposited with the American Documentation Institute. Order Document No. 5323 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance \$1.25 for 35 mm. microfilm or \$1.25 for 6 by 8 inch photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

<sup>3</sup> *Ibid.*

TABLE 1  
ITEMS RETAINED IN THE "SOCIAL ATTITUDES SCALE"

1. It is always very important to finish anything one has started.
2. At school, it is easy to find things to do when the teacher doesn't give us enough work.
3. Police cars should be especially marked so that you can always see them coming.
4. It is no use worrying about current events or public affairs; I cannot do anything about them anyhow.
5. We ought to worry about our own country and let the rest of the world take care of itself.
- <sup>a</sup>6. In school my behavior gets me into trouble.
- <sup>a</sup>7. I am hardly ever on time for meals.
- <sup>a</sup>8. I have been in trouble with the law or police.
9. When a person does not tell all his income in order to get out of paying some of his taxes, it is just as bad as stealing money from the government.
10. A person who does not vote when he can, is not a good citizen.
- <sup>a</sup>11. I hardly ever get my school work done on time.
- <sup>a</sup>12. I have played hookey from school.
13. Every citizen should take the time to find out about current events if it means giving up some spare time.
- <sup>a</sup>14. In school I am sometimes sent to the principal for being bad.
15. Maybe some minority groups (Negroes, Indians, Mexicans, Jews, etc.) do get bad treatment, but it's no business of mine.
16. We ought to let Europe get out of its own mess.
17. People criticize me for wasting time.
- <sup>a</sup>18. When I work on a committee, I usually let other people do most of the planning.
- <sup>a</sup>19. I am often late for school.
20. If it is worth starting at all, it is worth finishing.
- <sup>a</sup>21. I am the kind of person that people can count on.
- <sup>a</sup>22. In school, I am one of those who can go on working even though the teacher is out of the room.
23. People can count on me to get things done, without checking on me.
- <sup>a</sup>24. I am frequently chosen as a room helper or to run errands.
- <sup>a</sup>25. I do my chores the very best I know how.
- <sup>a</sup>26. I have been elected leader or president of my class.
27. Nothing is more important than to be honest with other people.
- <sup>a</sup>28. My teacher often complains because I don't finish my work.
29. When you can't do a job, it is no use to try to find someone else to do it.
30. It is more important to get the job done than worry about hurting other people's feelings.
31. Why bother to vote when you can do so little with just your one vote.
32. "Never give a sucker an even break."
33. Letting your friends down is not so bad because you can't do good all the time for everybody.
34. Our country would be a lot better off if we didn't have elections and people didn't have to vote.
35. It's a good thing the Atlantic Ocean separates us from Europe because then we don't have to worry about them.
36. It's more important to work for the good of the team than to work for your own good.
- <sup>a</sup>37. I would never let a friend down when he expects something of me.
38. People would be a lot better off if they could live far away from other people and never have to do anything for them.
39. Every person should give some of his time for the good of his town or city.
40. If everyone pitches in to do a job, it can always get done.
41. It is a good rule to do something for your neighbor if he does something for you.
42. Doing things which are important should come before things you enjoy doing.
43. When a person doesn't like something he is supposed to do he will try to get someone else to do it if he is smart.
44. Cheating on examinations is not so bad as long as nobody ever knows.
45. People have a real duty to take care of their parents when they are old, even if it costs a lot.
- <sup>a</sup>46. I usually work things out for myself rather than get someone to show me how.
- <sup>a</sup>47. I usually volunteer for special projects at school.
48. Children often get punished when they don't deserve it.
- <sup>a</sup>49. When given a task I stick to it even if things I like to do better come along.
50. It doesn't really matter whether parents attend Parent-Teacher meetings regularly.

<sup>a</sup> Designates "personal reference" item.

TABLE 2  
RESPONSIBILITY TEST SCORES FOR SEVERAL  
RURAL SAMPLES

Sample	Boys			Girls		
	N	$\bar{X}$	SD	N	$\bar{X}$	SD
10-year-olds, Nobles County	225	34.4	6.48	204	36.5	5.54
Conzemius' 10-year-olds						
Non-4-H members	85	35.8	4.51	92	38.5	5.48
4-H members	54	39.3	4.67	50	40.4	5.34
17-year-olds, Nobles County	119	37.8	5.96	112	41.6	4.43
Conzemius' high school seniors						
Non-4-H members	27	38.9	4.32	31	42.7	4.11
4-H members	34	41.1	4.67	29	43.6	3.19

bers in every comparison, though variances were not homogeneous in every instance. Whether this difference represents selection or the special effect of 4-H experience, she does not choose to say. Conzemius' experience is consistent with ours in finding greater sex differences than in our previous study, and a somewhat greater relationship of score to age than obtained in our previous work (3).

The "Social Attitudes Scale" yields a number of correlations with other measures which are of interest. Table 3 reports a number of these and shows that correlations of score with socioeconomic status are negligible, but with measures of adjustment they are moderately high.



The correlation of the new scale with the old scale on two different samples of children who took the two schedules four years apart gives values of  $+ .14$  and  $+ .32$ .

Whereas the "How I Feel About Citizenship" measure (1950) showed no age trend, a noticeable age relationship appears in the "Social Attitudes" test data (1954). From the youngest to the oldest age groups, the shift in mean score is approximately one standard deviation of the total distribution. Furthermore, the relationship to age is more pronounced in girls than in boys. The differences between the means of the sexes is about one-fourth a standard deviation of the combined distributions in the younger ages and more than one-half a standard deviation in the oldest age groups. In the sense of this test, girls are perhaps better "socialized" than boys.

It will be recalled that less than half the items of the 1954 schedule appeared in the 1950 measure. It may be that the 1950 measure, based strictly on empirically selected items, contained a large random component, and it is true that the reliability coefficients were not large (of the order of  $+ .64$  to  $+ .88$  for a two-week interval). The 1954 measure, designed for construct as well as concurrent validity, seems to have included more age-related items. The reliability coefficients were comparable, being of the order of  $+ .60$  to  $+ .70$  for a four-month retest interval. Both measures, however, showed correlations with adjustment measures of the order of  $+ .30$  to  $+ .40$  and better, and the "Social Attitudes Scale" consistently yields the larger values. Furthermore, there is a suggestion that the Social Attitudes test tends also in the adolescent period towards an increased association with other measures, especially measures of adjustment. There may be more organization or responsibility as a focal characteristic as the child matures in his experience.

In 1950, 26 of the 43 items (60%) were personal reference items, couched in the first person and having to do with the child's own conduct. In 1954, only 18 of the 50 items (36%) were of this character. The tests for a stratified sample<sup>6</sup> in age groups 10, 12, 14, and 16 were

<sup>6</sup> For other research purposes a "standard sample" of 50 boys and 50 girls had been selected at each year of age from our larger supply of children (running about 300 in each age group). This sample was stratified

TABLE 3  
CORRELATIONS OF RESPONSIBILITY TEST SCORES  
WITH CERTAIN OTHER VARIABLES: BOYS  
AND GIRLS COMBINED

Variable	Ages 9 & 10	Ages 11 & 12	Ages 13 & 14	Ages 15 & 16
Father's education	.15	.17	.24	.19
Father's occupation (1954)	.06	.07	.11	.01
"How I Feel About Things" <sup>a</sup>				
Adjustment score	.38	.44	.42	.44
Adjustment-maturity score	.34	.45	.46	.44
Sentence Completion Test score <sup>b</sup>	.03	.16	.23	.21

<sup>a</sup> A schedule in the Nobles County study of attitude and interest items selected empirically to discriminate between well adjusted and poorly adjusted children.

<sup>b</sup> This is also a Nobles County study instrument. It is scored according to the number of items showing positive affect in the completion supplied. This score has a low positive correlation with a criterion of adjustment.

TABLE 4  
MEANS AND STANDARD DEVIATIONS FOR TOTAL SCORE  
ON THE SOCIAL ATTITUDES MEASURE AND  
FOR PERSONAL REFERENCE ITEMS AND  
NONPERSONAL REFERENCE ITEMS

Age	Sex	N	Total Score		Personal Reference Items		Nonpersonal Reference Items	
			Mean	SD	Mean	SD	Mean	SD
10	B	93	34.61	5.02	11.98	2.60	22.63	4.42
	G	91	37.08	4.75	13.63	1.82	23.45	3.92
12	B	95	36.94	5.14	12.49	2.11	24.44	4.03
	G	98	39.42	3.99	13.86	2.23	25.57	4.14
14	B	92	36.48	6.77	11.36	2.73	25.12	5.04
	G	93	40.53	4.48	13.39	2.07	27.14	3.09
16	B	94	37.74	6.56	12.01	2.59	25.73	4.94
	G	91	41.57	3.97	13.43	2.01	28.14	3.92

recorded to provide separate scores on the personal reference items (see marked items in Table 1) and on the 32 nonpersonal reference items (unmarked items in Table 1.) Table 4 makes it clear that the age increment in score on the Social Attitudes measure as a whole is due to age change in the nonpersonal reference items.

An analysis of another aspect of the Nobles County work<sup>7</sup> has shown that adjustment items of the psychoneurotic type are also

according to occupation of the father, to represent the U.S. 1950 occupational distribution.

<sup>7</sup> As yet unpublished.

entirely of this personal reference type and are also free from an age relationship. Perhaps attitudes toward the community, society, and human relationships in general are modified by education, experience and maturity, while self-attitudes and personal habits are less amenable to change. Or, possibly, children progressively acquire more socially acceptable verbal statements about society and social relations as they move through school but fail to modify expressions about themselves, their self-feelings, and their own habits. Few studies of the "self image" have drawn this contrast, and this should prove a fruitful hypothesis for further work.

Correlations between the personal and non-personal subscores are modest, ranging from  $+ .29$  to  $+ .53$ , and being characteristically around  $+ .40$ .<sup>8</sup> There is no appreciable tendency for correlational values to change systematically with age, so that no argument can be offered concerning "integration" or "differentiation" of attitudes measured by these scales, over the age range represented in this study. Some information on the reliability of these subscales may be of interest. For a sample of 50 twelve-year-old girls the estimate of reliability obtained by Hoyt's analysis of variance method (5) was  $+ .73$ . For the subscale of 18 personal reference items, the estimate of reliability was  $+ .47$ , and for the 32 nonpersonal reference items the value was  $+ .63$ .

#### SUMMARY

A scale of social attitudes was designed to discriminate children who have, with their peers, a reputation for responsibility as con-

trasted with children who have little reputation for responsibility. The resulting scale yields a measure which is substantially correlated with other measures of personal and social adjustment. Unlike an earlier scale designed for a similar purpose, the present scale shows a positive trend in mean score with age. Analysis revealed that this age trend results from a larger proportion of items in the present scale which refers to social and societal obligations (or at least do not contain a highly personal reference as indicated by the first personal singular pronoun). Items which have a highly personal reference, referring to self-feelings, attitudes, and reports of personal behavior, do not show an age trend. The two sets of items are modestly intercorrelated, on the order of  $+ .40$ .

#### REFERENCES

1. CONZEMIUS, ROSEMARY. A comparison of the development of social responsibility in rural boys and girls. Unpublished master's paper, School of Social Work, Univer. of Minnesota, June, 1955.
2. GOUGH, H. G., McCLOSKEY, H., & MEEHL, P. E. A personality scale for social responsibility. *J. abnorm. soc. Psychol.*, 1952, **47**, 73-80.
3. HARRIS, D. B., CLARK, K. E., ROSE, A. M., & VALASEK, FRANCES. The measurement of responsibility in children. *Child Develpm.*, 1954, **25**, 21-28.
4. HARRIS, D. B., CLARK, K. E., ROSE, A. M., & VALASEK, FRANCES. The relationship of children's home duties to an attitude of responsibility. *Child Develpm.*, 1954, **25**, 29-33.
5. HOYT, C. J. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **3**, 153-60.
6. LAWSHE, C. H. A nomograph for estimating the validity of test items. *J. appl. Psychol.*, 1942, **26**, 846-849.

Received January 21, 1957.

<sup>8</sup> American Documentation Institute Document No. 5323, *op. cit.*



# A CROSS-CULTURAL SURVEY OF SOME SEX DIFFERENCES IN SOCIALIZATION<sup>1</sup>

HERBERT BARRY III, MARGARET K. BACON, AND IRVIN L. CHILD

*Yale University*

**I**N OUR society, certain differences may be observed between the typical personality characteristics of the two sexes. These sex differences in personality are generally believed to result in part from differences in the way boys and girls are reared. To the extent that personality differences between the sexes are thus of cultural rather than biological origin, they seem potentially susceptible to change. But how readily susceptible to change? In the differential rearing of the sexes does our society make an arbitrary imposition on an infinitely plastic biological base, or is this cultural imposition found uniformly in all societies as an adjustment to the real biological differences between the sexes? This paper reports one attempt to deal with this problem.

## DATA AND PROCEDURES

The data used were ethnographic reports, available in the anthropological literature, about socialization practices of various cultures. One hundred and ten cultures, mostly nonliterate, were studied. They were selected primarily in terms of the existence of adequate ethnographic reports of socialization practices and secondarily so as to obtain a wide and reasonably balanced geographical distribution. Various aspects of socialization of infants and children were rated on a 7-point scale by two judges (Mrs. Bacon and Mr. Barry). Where the ethnographic reports permitted, separate ratings were made for the socialization of boys and girls. Each rating was indicated as either confident or doubtful; with still greater uncertainty, or with complete lack of evidence, the particular rating was of course not made at all. We shall restrict the report of sex difference ratings to cases in which both judges made a confident rating. Also omitted is the one instance where the two judges reported a sex difference in opposite directions, as it demon-

strates only unreliability of judgment. The number of cultures that meet these criteria is much smaller than the total of 110; for the several variables to be considered, the number varies from 31 to 84.

The aspects of socialization on which ratings were made included:

1. Several criteria of attention and indulgence toward infants.

2. Strength of socialization from age 4 or 5 years until shortly before puberty, with respect to five systems of behavior; strength of socialization was defined as the combination of positive pressure (rewards for the behavior) plus negative pressure (punishments for lack of the behavior). The variables were:

(a) Responsibility or dutifulness training. (The data were such that training in the performance of chores in the productive or domestic economy was necessarily the principal source of information here; however, training in the performance of other duties was also taken into account when information was available.)

(b) Nurturance training, i.e., training the child to be nurturant or helpful toward younger siblings and other dependent people.

(c) Obedience training.

(d) Self-reliance training.

(e) Achievement training, i.e., training the child to orient his behavior toward standards of excellence in performance, and to seek to achieve as excellent a performance as possible.

Where the term "no sex difference" is used here, it may mean any of three things: (a) the judge found separate evidence about the training of boys and girls on this particular variable, and judged it to be identical; (b) the judge found a difference between the training of boys and girls, but not great enough for the sexes to be rated a whole point apart on a 7-point scale; (c) the judge found evidence only about the training of "children" on this variable, the ethnographer not reporting separately about boys and girls.

## SEX DIFFERENCES IN SOCIALIZATION

On the various aspects of attention and indulgence toward infants, the judges almost always agreed in finding no sex difference. Out of 96 cultures for which the ratings included the infancy period, 88 (92%) were rated with no sex difference by either judge for any of those variables. This result is consistent with the

<sup>1</sup> This research is part of a project for which financial support was provided by the Social Science Research Council and the Ford Foundation. We are greatly indebted to G. P. Murdock for supplying us with certain data, as indicated below, and to him and Thomas W. Maretzki for suggestions that have been used in this paper.

TABLE 1  
RATINGS OF CULTURES FOR SEX DIFFERENCES ON FIVE VARIABLES OF CHILDHOOD SOCIALIZATION PRESSURE

Variable	Number of Cultures	Both Judges Agree in Rating the Variable Higher in		One Judge Rates No Difference, One Rates the Variable Higher in		Percentage of Cultures with Evidence of Sex Difference in Direction of		
		Girls	Boys	Girls	Boys	Girls	Boys	Neither
Nurturance	33	17	0	10	0	82%	0%	18%
Obedience	69	6	0	18	2	35%	3%	62%
Responsibility	84	25	2	26	7	61%	11%	28%
Achievement	31	0	17	1	10	3%	87%	10%
Self-reliance	82	0	64	0	6	0%	85%	15%

point sometimes made by anthropologists that "baby" generally is a single status undifferentiated by sex, even though "boy" and "girl" are distinct statuses.

On the variables of childhood socialization, on the other hand, a rating of no sex difference by both judges was much less common. This finding of no sex difference varied in frequency from 10% of the cultures for the achievement variable up to 62% of the cultures for the obedience variable, as shown in the last column of Table 1. Where a sex difference is reported, by either one or both judges, the difference tends strongly to be in a particular direction, as shown in the earlier columns of the same table. Pressure toward nurturance, obedience, and responsibility is most often stronger for girls, whereas pressure toward achievement and self-reliance is most often stronger for boys.

For nurturance and for self-reliance, all the sex differences are in the same direction. For achievement there is only one exception to the usual direction of difference, and for obedience only two; but for responsibility there are nine. What do these exceptions mean? We have re-examined all these cases. In most of them, only one judge had rated the sexes as differently treated (sometimes one judge, sometimes the other), and in the majority of these cases both judges were now inclined to agree that there was no convincing evidence of a real difference. There were exceptions, however, especially in cases where a more formal or systematic training of boys seemed to imply greater pressure on them toward responsibility. The most convincing cases were the Masai and Swazi, where both judges had originally agreed in rating responsibility pressures greater in boys than in girls. In comparing the five aspects of socialization we may conclude that responsibility shows by far the strongest evidence of real variation in the direction of sex difference, and obedience

much the most frequently shows evidence of no sex difference at all.

In subsequent discussion we shall be assuming that the obtained sex differences in the socialization ratings reflect true sex differences in the cultural practices. We should consider here two other possible sources of these rated differences.

1. The ethnographers could have been biased in favor of seeing the same pattern of sex differences as in our culture. However, most anthropologists readily perceive and eagerly report novel and startling cultural features, so we may expect them to have reported unusual sex differences where they existed. The distinction between matrilineal and patrilineal, and between matrilocal and patrilocal cultures, given prominence in many ethnographic reports, shows an awareness of possible variations in the significance of sex differences from culture to culture.

2. The two judges could have expected to find in other cultures the sex roles which are familiar in our culture and inferred them from the material on the cultures. However, we have reported only confident ratings, and such a bias seems less likely here than for doubtful ratings. It might be argued, moreover, that bias has more opportunity in the cases ambiguous enough so that only one judge reported a sex difference, and less opportunity in the cases where the evidence is so clear that both judges agree. Yet in general, as may be seen in Table 1, the deviant cases are somewhat more frequent among the cultures where only one judge reported a sex difference.

The observed differences in the socialization of boys and girls are consistent with certain universal tendencies in the differentiation of adult sex role. In the economic sphere, men are more frequently allotted tasks that involve leaving home and engaging in activities where



a high level of skill yields important returns; hunting is a prime example. Emphasis on training in self-reliance and achievement for boys would function as preparation for such an economic role. Women, on the other hand, are more frequently allotted tasks at or near home that minister most immediately to the needs of others (such as cooking and water carrying); these activities have a nurturant character, and in their pursuit a responsible carrying out of established routines is likely to be more important than the development of an especially high order of skill. Thus training in nurturance, responsibility, and, less clearly, obedience, may contribute to preparation for this economic role. These consistencies with adult role go beyond the economic sphere, of course. Participation in warfare, as a male prerogative, calls for self-reliance and a high order of skill where survival or death is the immediate issue. The childbearing which is biologically assigned to women, and the child care which is socially assigned primarily to them, lead to nurturant behavior and often call for a more continuous responsibility than do the tasks carried out by men. Most of these distinctions in adult role are not inevitable, but the biological differences between the sexes strongly predispose the distinction of role, if made, to be in a uniform direction.<sup>2</sup>

The relevant biological sex differences are conspicuous in adulthood but generally not in childhood. If each generation were left entirely to its own devices, therefore, without even an older generation to copy, sex differences in role would presumably be almost absent in childhood and would have to be developed after puberty at the expense of considerable relearning on the part of one or both sexes. Hence, a pattern of child training which foreshadows adult differences can serve the useful function of minimizing what Benedict termed "discontinuities in cultural conditioning" (1).

The differences in socialization between the sexes in our society, then, are no arbitrary custom of our society, but a very widespread adaptation of culture to the biological substratum of human life.

#### VARIATIONS IN DEGREE OF SEX DIFFERENTIATION

While demonstrating near-universal tendencies in direction of difference between the socialization of boys and girls, our data do not show perfect uniformity. A study of the variations in our data may allow us to see some of the conditions which are associated with, and perhaps give rise to, a greater or smaller degree of this difference. For this purpose, we classified cultures as having relatively large or small sex difference by two different methods, one more inclusive and the other more selective. In both methods the ratings were at first considered separately for each of the five variables. A sex difference rating was made only if both judges made a rating on this variable and at least one judge's rating was confident.

In the more inclusive method the ratings were dichotomized, separately for each variable, as close as possible to the median into those showing a large and those showing a small sex difference. Thus, for each society a large or a small sex difference was recorded for each of the five variables on which a sex difference rating was available. A society was given an over-all classification of large or small sex difference if it had a sex difference rating on at least three variables and if a majority of these ratings agreed in being large, or agreed in being small. This method permitted classification of a large number of cultures, but the grounds for classification were capricious in many cases, as a difference of only one point in the rating of a single variable might change the over-all classification of sex difference for a culture from large to small.

In the more selective method, we again began by dichotomizing each variable as close as possible to the median; but a society was now classified as having a large or small sex difference on the variable only if it was at least one step away from the scores immediately adjacent to the median. Thus only the more decisive ratings of sex difference were used. A culture was classified as having an over-all large or small sex difference only if it was given a sex difference rating which met this criterion on at least two variables, and only if all such ratings agreed in being large, or agreed in being small.

We then tested the relation of each of these dichotomies to 24 aspects of culture on which

<sup>2</sup> For data and interpretations supporting various arguments of this paragraph, see Mead (2), Murdock (3), and Scheinfeld (6).

TABLE 2  
CULTURE VARIABLES CORRELATED WITH LARGE  
SEX DIFFERENCE IN SOCIALIZATION,  
SEPARATELY FOR TWO TYPES  
OF SAMPLE

Variable	More Selective Sample		More Inclusive Sample	
	$\phi$	N	$\phi$	N
Large animals are hunted	.48*	(34)	.28*	(72)
Grain rather than root crops are grown	.82**	(20)	.62**	(43)
Large or milking animals rather than small animals are kept	.65*	(19)	.43*	(35)
Fishing unimportant or absent	.42*	(31)	.19	(69)
Nomadic rather than sedentary residence	.61**	(34)	.15	(71)
Polygyny rather than monogamy	.51*	(28)	.38**	(64)

\*  $p < .05$ .

\*\*  $p < .01$ .

Note.—The variables have been so phrased that all correlations are positive. The phi coefficient is shown, and in parentheses, the number of cases on which the comparison was based. Significance level was determined by  $\chi^2$ , or Fisher's exact test where applicable, using in all cases a two-tailed test.

Murdock has categorized the customs of most of these societies<sup>3</sup> and which seemed of possible significance for sex differentiation. The aspects of culture covered include type of economy, residence pattern, marriage and incest rules, political integration, and social organization. For each aspect of culture, we grouped Murdock's categories to make a dichotomous contrast (sometimes omitting certain categories as irrelevant to the contrast). In the case of some aspects of culture, two or more separate contrasts were made (e.g., under form of marriage we contrasted monogamy with polygyny, and also contrasted sororal with nonsororal polygyny). For each of 40 comparisons thus formed, we prepared a 2 x 2 frequency table to determine relation to each of our sex-difference dichotomies. A significant relation was found for six of these 40 aspects of culture with the more selective dichotomization of over-all sex difference. In four of these comparisons, the relation to the more inclusive dichotomization was also significant. These relationships are all given in Table 2, in the form of phi coefficients, along with the outcome of testing significance by the use of  $\chi^2$  or Fisher's exact test. In trying to interpret these findings, we have also considered the nonsignificant correlations

<sup>3</sup> These data were supplied to us directly by Professor Murdock.

with other variables, looking for consistency and inconsistency with the general implications of the significant findings. We have arrived at the following formulation of results:

1. Large sex difference in socialization is associated with an economy that places a high premium on the superior strength, and superior development of motor skills requiring strength, which characterize the male. Four of the correlations reported in Table 2 clearly point to this generalization: the correlations of large sex difference with the hunting of large animals, with grain rather than root crops, with the keeping of large rather than small domestic animals, and with nomadic rather than sedentary residence. The correlation with the unimportance of fishing may also be consistent with this generalization, but the argument is not clear.<sup>4</sup> Other correlations consistent with the generalization, though not statistically significant, are with large game hunting rather than gathering, with the hunting of large game rather than small game, and with the general importance of all hunting and gathering.

2. Large sex difference in socialization appears to be correlated with customs that make for a large family group with high cooperative interaction. The only statistically significant correlation relevant here is that with polygyny rather than monogamy. This generalization is, however, supported by several substantial correlations that fall only a little short of being statistically significant. One of these is a

<sup>4</sup> Looking (with the more inclusive sample) into the possibility that this correlation might result from the correlation between fishing and sedentary residence, a complicated interaction between these variables was found. The correlation of sex differentiation with absence of fishing is found only in nomadic societies, where fishing is likely to involve cooperative activity of the two sexes, and its absence is likely to mean dependence upon the male for large game hunting or herding large animals (whereas in sedentary societies the alternatives to fishing do not so uniformly require special emphasis on male strength). The correlation of sex differentiation with nomadism is found only in nonfishing societies; here nomadism is likely to imply large game hunting or herding large animals, whereas in fishing societies nomadism evidently implies no such special dependence upon male strength. Maximum sex differentiation is found in nomadic nonfishing societies (15 with large difference and only 2 with small) and minimum sex differentiation in nomadic fishing societies (2 with large difference and 7 with small difference). These findings further strengthen the argument for a conspicuous influence of the economy upon sex differentiation.



correlation with sororal rather than nonsororal polygyny; Murdock and Whiting (4) have presented indirect evidence that co-wives generally shows smoother cooperative interaction if they are sisters. Correlations are also found with the presence of either an extended or a polygynous family rather than the nuclear family only; with the presence of an extended family; and with the extreme contrast between maximal extension and no extension of the family. The generalization is also to some extent supported by small correlations with wide extension of incest taboos, if we may presume that an incest taboo makes for effective unthreatening cooperation within the extended family. The only possible exception to this generalization, among substantial correlations, is a near-significant correlation with an extended or polygynous family's occupying a cluster of dwellings rather than a single dwelling.<sup>6</sup>

In seeking to understand this second generalization, we feel that the degree of social isolation of the nuclear family may perhaps be the crucial underlying variable. To the extent that the nuclear family must stand alone, the man must be prepared to take the woman's role when she is absent or incapacitated, and vice versa. Thus the sex differentiation cannot afford to be too great. But to the extent that the nuclear family is steadily interdependent with other nuclear families, the female role in the household economy can be temporarily taken over by another woman, or the male role by another man, so that sharp differentiation of sex role is no handicap.

The first generalization, which concerns the economy, cannot be viewed as dealing with material completely independent of the ratings of socialization. The training of children in

<sup>6</sup> We think the reverse of this correlation would be more consistent with our generalization here. But perhaps it may reasonably be argued that the various nuclear families composing an extended or polygynous family are less likely to develop antagonisms which hinder cooperation if they are able to maintain some physical separation. On the other hand, this variable may be more relevant to the first generalization than to the second. Occupation of a cluster of dwellings is highly correlated with presence of herding and with herding of large rather than small animals, and these economic variables in turn are correlated with large sex difference in socialization. Occupation of a cluster of dwellings is also correlated with polygyny rather than monogamy and shows no correlation with sororal vs. nonsororal polygyny.

their economic role was often an important part of the data used in rating socialization variables, and would naturally vary according to the general economy of the society. We would stress, however, that we were by no means using the identical data on the two sides of our comparison; we were on the one hand judging data on the socialization of children and on the other hand using Murdock's judgments on the economy of the adult culture. In the case of the second generalization, it seems to us that there was little opportunity for information on family and social structure to have influenced the judges in making the socialization ratings.

Both of these generalizations contribute to understanding the social background of the relatively small difference in socialization of boys and girls which we believe characterizes our society at the present time. Our mechanized economy is perhaps less dependent than any previous economy upon the superior average strength of the male. The nuclear family in our society is often so isolated that husband and wife must each be prepared at times to take over or help in the household tasks normally assigned to the other. It is also significant that the conditions favoring low sex differentiation appear to be more characteristic of the upper segments of our society, in socioeconomic and educational status, than of lower segments. This observation may be relevant to the tendency toward smaller sex differences in personality in higher status groups (cf. Terman and Miles, 8).

The increase in our society of conditions favoring small sex difference has led some people to advocate a virtual elimination of sex differences in socialization. This course seems likely to be dysfunctional even in our society. Parsons, Bales, *et al.* (5) argue that a differentiation of role similar to the universal pattern of sex difference is an important and perhaps inevitable development in any social group, such as the nuclear family. If we add to their argument the point that biological differences between the sexes make most appropriate the usual division of those roles between the sexes, we have compelling reasons to expect that the decrease in differentiation of adult sex role will not continue to the vanishing point. In our training of children, there may now be less differentiation in sex role than characterizes

adult life—so little, indeed, as to provide inadequate preparation for adulthood. This state of affairs is likely to be especially true of formal education, which is more subject to conscious influence by an ideology than is informal socialization at home. With child training being more oriented toward the male than the female role in adulthood, many of the adjustment problems of women in our society today may be partly traced to conflicts growing out of inadequate childhood preparation for their adult role. This argument is nicely supported in extreme form by Spiro's analysis of sex roles in an Israeli kibbutz (7). The ideology of the founders of the kibbutz included the objective of greatly reducing differences in sex role. But the economy of the kibbutz is a largely nonmechanized one in which the superior average strength of men is badly needed in many jobs. The result is that, despite the ideology and many attempts to implement it, women continue to be assigned primarily to traditional "women's work," and the incompatibility between upbringing or ideology and adult role is an important source of conflict for women.

*Note on regional distribution.* There is marked variation among regions of the world in typical size of sex difference in socialization. In our sample, societies in North America and Africa tend to have large sex difference, and societies in Oceania to have small sex difference. Less confidently, because of the smaller number of cases, we can report a tendency toward small sex differences in Asia and South America as well. Since most of the variables with which we find the sex difference to be significantly correlated have a similar regional distribution, the question arises whether the correlations might better be ascribed to some quite different source having to do with large regional similarities, rather than to the functional dependence we have suggested. As a partial check, we have tried to determine whether the correlations we report in Table 2 tend also to be found strictly within regions. For each of the three regions for which we have sizable samples (North America, Africa, and Oceania) we have separately plotted 2 x 2 tables corresponding to each of the 6 relationships reported in Table 2. (We did this only for the more inclusive sample, since for the more selective sample the

number of cases within a region would have been extremely small.) Out of the 18 correlations thus determined, 11 are positive and only 3 are negative (the other 4 being exactly zero). This result clearly suggests a general tendency for these correlations to hold true within regions as well as between regions, and may lend further support to our functional interpretation.

### SUMMARY

A survey of certain aspects of socialization in 110 cultures shows that differentiation of the sexes is unimportant in infancy, but that in childhood there is, as in our society, a widespread pattern of greater pressure toward nurturance, obedience, and responsibility in girls, and toward self-reliance and achievement striving in boys. There are a few reversals of sex difference, and many instances of no detectable sex difference; these facts tend to confirm the cultural rather than directly biological nature of the differences. Cultures vary in the degree to which these differentiations are made; correlational analysis suggests some of the social conditions influencing these variations, and helps in understanding why our society has relatively small sex differentiation.

### REFERENCES

1. BENEDICT, RUTH. Continuities and discontinuities in cultural conditioning. *Psychiatry*, 1938, 1, 161-167.
2. MEAD, MARGARET. *Male and female*. New York: Morrow, 1949.
3. MURDOCK, G. P. Comparative data on the division of labor by sex. *Social Forces*, 1937, 15, 551-553.
4. MURDOCK, G. P., & WHITING, J. W. M. Cultural determination of parental attitudes: The relationship between the social structure, particularly family structure and parental behavior. In M. J. E. Senn (Ed.), *Problems of infancy and childhood: Transactions of the Fourth Conference*, March 6-7, 1950. New York: Josiah Macy, Jr. Foundation, 1951. Pp. 13-34.
5. PARSONS, T., BALES, R. F., et al. *Family, socialization and interaction process*. Glencoe, Ill.: Free Press, 1955.
6. SCHEINFELD, A. *Women and men*. New York: Harcourt, Brace, 1944.
7. SPIRO, M. E. *Kibbutz: Venture in Utopia*. Cambridge: Harvard Univer. Press, 1956.
8. Terman, L. M., & MILES, CATHERINE C. *Sex and personality*. New York: McGraw-Hill, 1936.

Received February 6, 1957.



# AVOIDANCE LEARNING OF PERCEPTUAL DEFENSE AND VIGILANCE<sup>1</sup>

DON E. DULANY, JR.

*University of Illinois*

THOSE who argue for a perceptual defense mechanism analogous to repression and those who believe that the adaptive value of sensitization to threat suggests a vigilance mechanism can both find support in a wealth of controversial data (3, 7)—as well as can those who would deny both views (4, 11, 12, 13). Much of the confusion has resulted from the use of linguistic stimuli whose significance to the subject is a matter of dispute. So long as the history of the subjects' commerce with the stimulus materials remains uncertain, convenient assumptions permit any of the several views to be salvaged in the face of most data. The chief object of the present study is to produce experimentally the learning of perceptual defense and vigilance, and in so doing, to relate the mechanisms to antecedent experimental procedures.

Knowing that a stimulus is anxiety arousing does not in itself permit us to predict the perceptual reaction to that stimulus. Lazarus, Eriksen, and Fonda (6) have demonstrated both perceptual defense and vigilance reactions to the same hostile or sexual, and presumably threatening, material; which reaction occurs is predicted from diagnostic categories. Encountering stimuli with sexual or hostile meaning is no novelty to the adult subject, who has had ample opportunity for learning to handle any anxiety they engender. We might guess that one kind of subject has learned to defend against threatening stimuli because with his particular experiences and personal economy that reaction has somehow been to his advantage. Another has learned perceptual vigilance because that reaction served him well. The specific learning mechanism involved has not been made clear. Hence, a second purpose of this study is to offer an account in terms of behavior theory of how perceptual defense and vigilance may be learned and to test two hypotheses derived from that account.

This analysis extends to perception a two-stage conception of avoidance learning that is essentially in agreement with the views of Mowrer (9) and of Solomon (16). Once an anxiety response has been classically conditioned to a previously neutral stimulus, any number of perceptual reactions are theoretically possible—vigilance and defense, as well as many qualitative distortions. The convenience of certain psychophysical methods has led to great experimental interest in two such reactions along the single dimension of recognizability. We may think of these potential perceptual reactions to a threatening stimulus as ordered in a hierarchy of probability of occurrence. There is some evidence that perceptual vigilance holds the dominant position in an initial hierarchy of perceptual reactions to threat. Pustell (14) produced perceptual vigilance experimentally by classically associating an electric shock with a number of neutral geometric figures. Lazarus and McCleary (5) classically conditioned GSRs to nonsense syllables, obtaining their threshold measures under threat of further punishment, but failed to find a significant difference between thresholds for their shock syllables and a number of nonshock syllables. Could a reluctance to name the shock syllable have cancelled out a primitive vigilance reaction? Lysak (8) replicated their finding, then in another experiment removed the electrodes before obtaining threshold measures and found thresholds for his shock syllables to be lower than those for nonshock syllables. These findings suggest that perceptual vigilance is dominant in an initial hierarchy of perceptual reactions to a threatening stimulus, although response suppression may obscure the effect when there is a continuing imminent possibility of punishment. The superior adaptive value of such a mechanism is obvious.

Whatever perceptual reaction is initially dominant, it should be possible to realign this hierarchy by introducing selective reinforcement. The perceptual reaction that is instrumental to the avoidance of punishment and reduction of the conditioned anxiety should be

<sup>1</sup>This article is based on a doctoral dissertation submitted to the Department of Psychology, University of Michigan. I wish to express my gratitude to Dr. E. L. Walker for his generous advice and guidance as chairman of my doctoral committee.

strengthened. In the course of ordinary experience, perceptual defense and vigilance occur in the presence of varied and competing stimuli. They name two kinds of imbalance among a set of competing perceptual responses. If one perceptual response is punished while competing perceptual responses are instrumental to the avoidance of punishment and a reduction in anxiety, perceptual defense should be learned, since competing perceptual responses should tend to crowd out the critical perceptual response. On the other hand, perceptual vigilance should be learned when one perceptual response is instrumental to avoidance of punishment while competing perceptual responses are punished. Specifically, the following hypotheses are tested: (a) When one perceptual response is followed by punishment and competing perceptual responses are instrumental to avoidance of punishment, the punished response becomes weaker as compared with the competing perceptual responses; (b) When one perceptual response is instrumental to avoidance of punishment and competing perceptual responses are punished, the avoidance response becomes stronger as compared with the competing perceptual responses.

## METHOD

### Subjects

The *Ss* were 32 undergraduates at the University of Michigan, evenly divided between the defense and vigilance training procedures. They were randomly assigned to the two procedures, the only requirement being that an equal number of men and women participate in each. All *Ss* were volunteers, and none was aware of the purpose of the experiment.

### Stimulus Materials

Four rather meaningless, emotionally neutral figures—a circle, diamond, square, and triangle—were cut from black Munsell paper and mounted on white cards. The figures were arranged top, bottom, left, and right, on each of four cards, so that each figure appeared in all four positions. Hence, if each of the four cards was presented to the *S* an equal number of times, each figure would appear an equal number of times in each of the four positions. In this way, positional (see 10) and frequency effects were controlled. Prior to this experiment, the figures had been roughly equated for discriminability by Thomas Pustell.

### Apparatus

A Gerbrands tachistoscope, designed for individual use, was adjusted for a constant exposure interval of .12 sec. throughout all experimental series. With a pre-exposure field of 6.29 ft. lamberts, an exposure

field of 0.425 ft. lamberts, and geometrical forms of less than one inch diameter, this exposure time produced a level of awareness too low for the figures to be readily named.

The electric shock applicator contained four pairs of electrodes spaced about an inch apart and was designed to be strapped to the leg of the *S* just above the ankle. In order to discourage adaptation and at the same time prevent burns, the shock could be sent through first one pair of electrodes and then another from trial to trial. The shock circuits and applicator were designed by Dr. Carl Brown.

### Defense Training Experiment

*Level of awareness series.* The experimental session began for each *S* with an assessment of his level of awareness. A particular exposure duration, together with illumination level, does not define a single level of awareness for more than one *S*. To ascertain whether level of awareness was related to the amount and character of the experimental effect, the *S* was given a simple discrimination task. A forced-choice technique developed by Blackwell (1) and modified by Blum (2) was selected for its sensitivity in measuring discrimination without awareness. It has the additional value of permitting a number of stimuli to be presented at once, a condition that more truly represents most perceptual situations.

The *S* was informed that he would be shown a number of cards, but that they would be flashed too briefly for him to see much. On each trial, *E* asked *S* to "Find the square" or "Find the diamond," etc. The *S* responded with "Top," "Bottom," "Left," "Right," depending on where he thought he saw the figure. He was instructed on each trial to focus on the fixation point in the center of the field. Sixty-four trials were run, and *S*'s level of awareness was measured by his percentage of correct identifications.

*Baseline series.* Although the figures were roughly equal in recognizability for a group collectively, individuals find one figure or another more salient because of their peculiar perceptual histories and predilections. For this reason an empirical baseline of recognizability of each figure was obtained for every *S*. The procedure for this series was like that of the preceding in every way except that the *S* was instructed to "tell me on each trial which figure is *clearest*—which one is most *recognizable*." He was again asked to convey his choice by saying "Top," "Bottom," "Left," or "Right," and to follow his hunch whenever uncertain. Any experimental effect produced



by the following training procedure should register as a departure from this empirical baseline.

*Training series.* The defense training procedure was essentially like the baseline series, with the important difference that reinforcement was introduced. One of the geometrical figures, selected arbitrarily, was designated the critical stimulus. Each of the figures was the critical one for four of the 16 Ss. As in the preceding series, S was instructed to report on each trial which of the four figures was most recognizable by indicating its position. When the critical stimulus was selected, an electric shock followed two seconds after exposure of the figures and lasted another two seconds. Selection of one of the other three figures as most recognizable resulted in avoidance of the shock.

Circuits were designed so as to permit S's pressing of one of four buttons to reveal his choice of figures and control presentation of the shock. The critical stimulus, of course, varied in position from trial to trial, and therefore no single one of the four responses was always punished. The E noted the position of the critical stimulus before each trial, and adjusted a commutator, thereby determining which of the four buttons would activate the shock circuit.

For the purposes of this experiment it was imperative that the shock be punishing. It was necessary that it be decidedly unpleasant, even painful for best results, but short of incapacitating. Since individual pain thresholds for electric shock vary so widely, the intensity of shock was set at a subjective level for each S. After he had been assured that the shock was quite harmless and encouraged to take as much as he could stand, the shock was set at the level each S judged to be "just this side of beyond endurance." By phenomenological consensus of 32 Ss, the shock was decidedly unpleasant, even painful, and like the jab of a hot needle.

Sixty-four trials were run, approximately one every 10 seconds. After each block of 16 trials, S rested briefly.

*Assessment series.* Immediately after the training series, S was allowed to remove the shock applicator and was assured that there would be no more shock. The instructions and procedure for the assessment series were identical with those for the baseline series. The same cards were presented for 64 trials, and on

each trial S indicated which figure was most recognizable. Finally, he was questioned for any knowledge of the rule by which he had been shocked.

### *Vigilance Training Experiment*

The vigilance training experiment repeated the defense training experiment with one important difference: the reinforcement prescription was reversed. When S selected one of the three noncritical stimuli as most recognizable, he was shocked; when he selected the critical stimulus, he avoided the shock.

### RESULTS AND DISCUSSION

Relative perceptual response strength for each figure was measured by the frequency with which that figure was judged clearer than competing stimuli. The Ss of the defense training procedure should, if the first hypothesis is correct, select the critical figure less frequently in the assessment series, after training, than they did in the baseline series. Amount of shift in frequency of response ranged from 0 to -14 with a mean of -5.8. In order to evaluate the group shift, it was noted that 14 of the 16 Ss shifted in the predicted direction while two remained unchanged. By the sign test this shift is significant with a  $p < .01$ .<sup>2</sup> These findings confirm Hypothesis I.

Hypothesis II leads to the contrary expectation that Ss of the vigilance training procedure should select the critical stimulus more frequently during the assessment series than during the baseline series. Magnitude of shift ranged from -7 to +36 with a mean of +7.0. Thirteen of 16 Ss shifted in the anticipated direction, and this shift is significant by the sign test with a  $p < .025$ . Thus the second hypothesis is supported.

We would also like to know what progressive changes take place throughout the training series, whether these changes follow a typical learning curve, and whether there is any sign of extinction in the assessment series. Consequently, pooled frequency of selection of the critical stimulus was graphed as a function of trials.

For the defense training group (see Fig. 1), frequency of selection drops precipitously from the baseline to the early part of the training series. Apparently the learning takes place very rapidly, although the abruptness of the drop

<sup>2</sup> Since the direction of experimental shift was predicted, one-tail  $p$  values are reported.

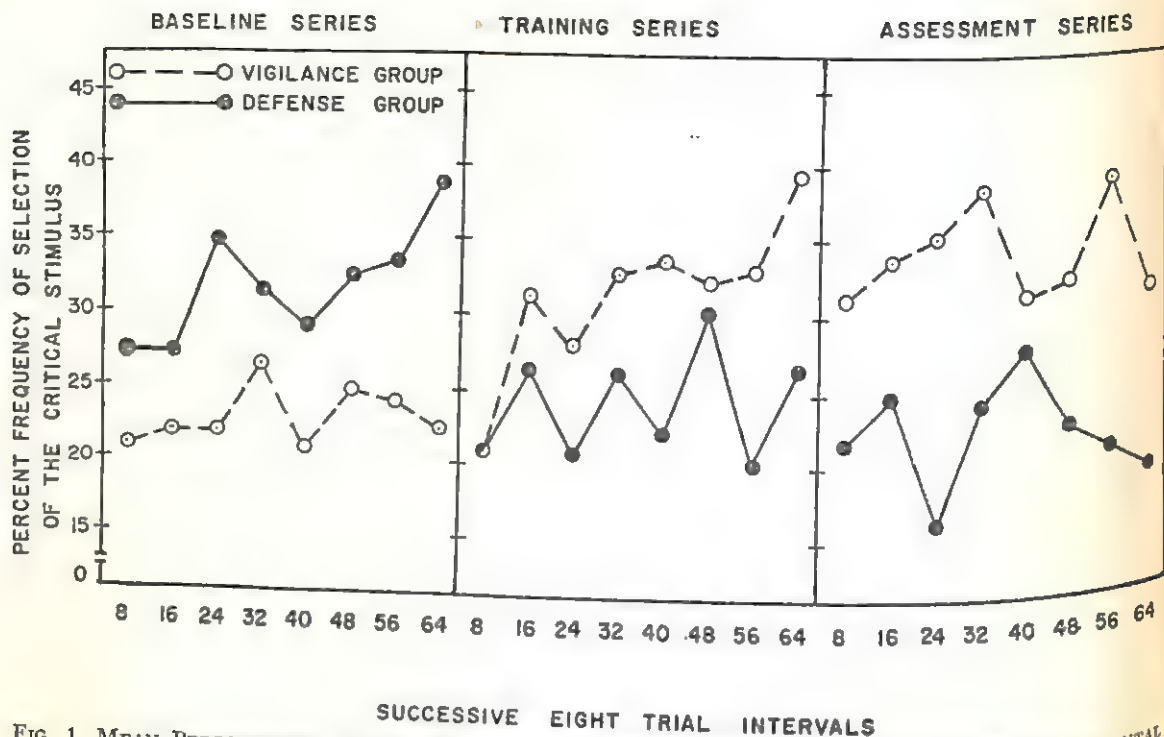


FIG. 1. MEAN PERFORMANCE OF THE DEFENSE AND VIGILANCE GROUPS THROUGHOUT EACH EXPERIMENTAL SERIES

may be partly due to a chance increase in frequency toward the end of the baseline series. In the vigilance training series, frequency of selection increases throughout the training series in a manner strongly suggestive of a learning curve. Since an anxiety reaction is very probably conditioned early in both training series, or even earlier, when the shock is adjusted to the appropriate subjective level, there is some reason to expect the instrumental learning of defense to proceed more rapidly than the learning of vigilance. Theoretically, vigilance is acquired as a single, critical, perceptual response, strengthened by anxiety reduction attendant upon shock avoidance; this is the kind of learning that ought to be describable by a familiar learning curve. The learning of perceptual defense, however, proceeds as the three competing perceptual responses are strengthened, by anxiety reduction, at the expense of the critical perceptual response. There is less basis for anticipating the function for this kind of learning. That the learning of defense did occur more rapidly than the learning of vigilance might be explained by the greater probability of reinforcement in the defense series. It is also likely that the greater number of shocks in the vigilance series

produced a higher level of anxiety in the vigilance group. Conceivably, level of anxiety drive could be related to the rapidity and course of learning. Finally, for neither procedure is there any evidence of extinction in the assessment series; and in view of the well-documented resistance to extinction of avoidance responses (15, 17) this is not too surprising.

A problem of interpretation arises from the fact that the empirical baseline values differ somewhat from their expected value of 25%—31.6% for the defense group and 23.0% for the vigilance group. While there is no reasonable alternative but to attribute this disparity to sampling fluctuation, it is important to distinguish two possible sources of error. Have two deviant samples been drawn from a population of responses that mark the baseline? Or have such deviant samples been drawn from a population of Ss varying in perceptual sensitivity to the critical stimuli? In the former case, we should expect some regression toward the value of 25% in the assessment series, quite apart from any effect of the training procedures. If the latter possibility is the case, then there is no problem, and the experimental effect is the more impressive for having overridden figure preferences. It is



possible, for purposes of analysis, to eliminate those four *Ss* with the highest frequencies at the defense baseline and the three *Ss* lowest at the vigilance baseline, leaving the two baseline means approximately equal. Of the 12 *Ss* remaining in the defense group, 10 selected the critical stimulus less frequently during the assessment series than at the baseline. By the sign test this shift is significant with  $p < .025$ . Ten of the 13 *Ss* remaining in the vigilance group shifted in the predicted direction ( $p < .05$ ). It seems clear that the regression hypothesis cannot account for these results.

How justified is the interpretation that these training procedures have changed the strength of perceptual responses rather than *S*'s disposition to make a report that has invariably brought him a painful electric shock? If *S* at any time discerned a connection between pressing the button for a square and being shocked, it would be only prudent for him to take a good look on each trial, then press one of the other buttons. For a number of reasons this suggested process cannot account for these results. Only two of 32 *Ss* could name the critical figure or verbalize the rule by which they had been shocked. Typical of their replies when questioned were, "Maybe there was a pattern to it, but I don't know what it was" and "I didn't know what I did or saw would have anything to do with it." This is, in fact, learning without awareness. Moreover, if the *S* were deliberately falsifying his report, then during the assessment series, after the shock applicator had been removed, he should have abandoned this tactic; frequency of selection of the critical stimulus would have returned to what it was at the baseline. It is evident from the data that this did not happen. Furthermore, the greater his awareness, the more information he has to act on and the more successful he should be; yet no relation was found between level of awareness and magnitude of experimental effect.

If the theoretical assumptions of the study are correct, something like the following occurred. An anxiety reaction produced by the shock was conditioned to the stimulus figures. It seems likely that anxiety attached to all four figures rather than just to the critical figure because of their low discriminability at a low level of awareness. Upon presentation of the stimulus, *S* tensed, gasped, or wrapped a leg around the chair, regardless of which figure

was selected. The total conditioned stimulus was probably even more complex, consisting of a stimulus presentation plus a lapse of time of two seconds. Generally, anxiety reduction is assumed to follow upon withdrawal of the cues eliciting the anxiety. However, termination of the exposure period brought no visible end to the *S*'s malaise. It was not until two seconds had elapsed and no shock had occurred that signs of relaxation or a relieved and grateful sigh appeared. In the absence of any good measure of momentary changes in the anxiety level it is reasonable to think anxiety drive reduction and reinforcement of the prior perceptual response occurred at this point.

This interpretation, which holds that the punished response is "weakened" only because competing responses are strengthened is advanced because it has worked well in describing the avoidance learning of skeletal responses (9, 16). Logically, of course, the shift in relative perceptibility of these figures could be credited either to sensitization to the nonpunished stimuli or desensitization to the punished stimuli, or both. The data prove only that at least one of these reactions must have occurred. This ambiguity is shared, however, with studies in which a single stimulus is presented by the ascending method of limits. In studies following the latter design, the assessed strength of the veridical percept is relative to the strength of competing perceptual hypotheses (12). The forced-choice procedure simply delimits the range of competing perceptual responses to a high priority few. Until the evidence for sensitization and desensitization is less equivocal, perceptual vigilance and defense can legitimately be identified only with shifts in balance among competing perceptual responses, in the one case toward a critical percept, and in the other case against it.

This study supports the view that perceptual defense and vigilance are learned reactions to anxiety arousing stimuli. It is consistent with Postman's recent statement that "the facts which give rise to this concept [of perceptual defense] can better be subsumed under other, more general, principles of perception" (12, p. 298); the same could be said for vigilance. Both concepts, it is suggested, may be subsumed under more general principles of perceptual learning, but this statement is not to be construed as asserting that perceptual defense and vigilance are therefore explained away.

Though the mechanisms appear to follow more general principles of learning, they nevertheless retain some identity as mechanisms.

### SUMMARY

This study experimentally produces the learning of perceptual defense and vigilance and proposes a behavior theory analysis of the learning process. According to this analysis, perceptual defense is learned when the perceptual response to a threatening stimulus is punished and competing perceptual responses are instrumental to anxiety reduction. Competing perceptual responses when reinforced are strengthened at the expense of the critical perceptual response. Perceptual vigilance is learned when the perceptual response to a threatening stimulus is reinforced by anxiety reduction and competing perceptual responses are punished.

Avoidance learning procedures were employed to change the comparative recognizability of four geometrical figures presented simultaneously below the threshold for awareness. Sixteen Ss underwent defense training, another 16 vigilance training. Before each training series, the comparative recognizability of the four figures was assessed for each S by a forced-choice technique. In both training series, S reported on each trial the position of the figure he found clearest. During defense training, if he selected the critical figure, he was shocked. If he selected one of the other three figures, he avoided the shock. During vigilance training, selection of the critical figure was instrumental to shock avoidance while selection of any of the other three figures incurred the shock. After each training procedure, comparative recognizability of the figures was assessed again. Fourteen of 16 defense Ss found the critical figure less recognizable after training than before. Thirteen of 16 vigilance Ss found the critical figure more recognizable than before training. Learning in both cases proceeded in the absence of awareness.

### REFERENCES

1. BLACKWELL, R. Psychophysical thresholds: experimental studies of methods of measurement. *Eng. res. Bull. No. 36*, Univer. of Michigan, 1953.
2. BLUM, G. S. An experimental reunion of psychoanalytic theory with perceptual vigilance and defense. *J. abnorm. soc. Psychol.*, 1954, **13**, 94-99.
3. ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175-182.
4. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401-410.
5. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58**, 113-123.
6. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471-482.
7. LAZARUS, R. S. Is there a mechanism of perceptual defense? A reply to Postman, Bronson, and Gropper. *J. abnorm. soc. Psychol.*, 1954, **49**, 396-398.
8. LYSAK, W. The effects of punishment upon syllable recognition thresholds. *J. exp. Psychol.*, 1954, **47**, 343-351.
9. MOWLER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem solving." *Harvard educ. Rev.*, 1944, Spring, 102-148.
10. OHRBACH, J. Retinal locus as a factor in recognition. *Amer. J. Psychol.*, 1952, **65**, 555-562.
11. POSTMAN, L. On the problem of perceptual defense. *Psychol. Rev.* 1951, **58**, 113-123.
12. POSTMAN, L. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation: A symposium*. Lincoln: Univer. of Nebraska Press, 1953.
13. POSTMAN, L., BRONSON, W. C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215-225.
14. PUSTELL, T. Cue and drive aspects of anxiety in relation to perceptual vigilance and defense. Unpublished doctoral dissertation, Univer. of Michigan, 1955.
15. SIDMAN, M. On the persistence of avoidance behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 217-221.
16. SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: Acquisition in normal dogs. *Psychol. Monogr.* 1953, **67**, No. 4, (Whole No. 354).
17. SOLOMON, R. L., KAMIN, L. J., & WYNNE, L. C. Traumatic avoidance learning: the outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, **48**, 206-216.

Received February 11, 1957.



# THE RELATIVE AVERSIVENESS OF WARNING SIGNAL AND SHOCK IN AN AVOIDANCE SITUATION

MURRAY SIDMAN AND JOHN J. BOREN<sup>1</sup>

*Walter Reed Army Institute of Research*

## EXPERIMENT I

### *Subjects and Apparatus*

The Ss were 4 male albino rats. The lever-pressing apparatus, with its associated automatic programming and recording devices, has been described previously (12, 14). In brief, the experimental chamber was a metal box with a grid floor and a sloping plexiglass ceiling. Depression of a lever, which was inserted through one wall of the chamber, constituted the measured avoidance response. A small neon pilot light was employed as the stimulus. The behavior was recorded on magnetic counters and cumulative recorders.

### *Procedure*

*Avoidance conditioning.* The avoidance response (lever-pressing) was first conditioned by means of a previously described "nondiscriminated" avoidance procedure (11). Brief shocks, of a fixed duration (approximately 0.3 sec.), were delivered through the grid floor every 20 seconds. Each time the animal depressed the lever, however, the shock delivery was postponed an additional 20 seconds, measured from the lever depression. Only the downward movement of the lever postponed the shock; continued holding of the lever did not affect the time of shock delivery.

The interval by which each response delays the shock has been termed the response-shock (RS) interval. A response-shock interval of 20 seconds, with no stimulus light present, is referred to as (RS20")<sup>D</sup>. The superscript "D" indicates the absence of the stimulus light.

*Chained avoidance contingencies under stimulus control.* After the rate of avoidance responding became reasonably stable on the shock schedule, (RS20")<sup>D</sup>, the following procedure was instituted. If the animal permitted 15 seconds to elapse without a lever-pressing response, a stimulus light (L) would come on. Every lever press in the absence of the stimulus, however, postponed the onset of L for an additional 15 seconds. In the absence of L, then, there is a response-light (RL) interval instead of a response-shock (RS) interval. This is referred to as (RL15")<sup>D</sup>, to indicate that each response in the absence of the light postponed the onset of the light for 15 seconds.

When the animal permitted the light to come on, a different schedule went into effect. If there was no response in the presence of the light a shock occurred within five seconds and the light terminated, reinstating (RL15")<sup>D</sup>. Each response in the light, however, postponed the shock for five seconds. The light would remain on until the animal permitted the shock to occur. Thus, in the presence of the light there is an RS interval of five seconds, which may be denoted as (RSS")<sup>L</sup>.

It is a widespread clinical and personal observation that the anticipation of pain may be worse than the pain itself. Yet the phenomenon has received little experimental attention. Possibly relevant observations are those of Gwinn (2) and Solomon, Kamin, and Wynne (18), who found that extinction of avoidance behavior was not hastened when the avoidance response was punished by shock. Kamin (4) demonstrated that a response which terminates a warning stimulus can be conditioned even when the response is regularly shocked. Escape from the warning stimuli was possibly sufficiently reinforcing to counteract the effects of punishment. Hill, Kornetsky, Flanary, and Wikler (3) found that shock intensities were overestimated by subjects who were not familiarized and reassured with respect to "the potentially fear-inspiring experimental situation" prior to the experiment.

A more direct experimental demonstration would consist of an arrangement that permits objective measurement of the subject's preference, either for a noxious event or for an anticipatory situation prior to the noxious event. The present study demonstrates one such arrangement in the framework of a modified avoidance conditioning procedure. In the classical avoidance experiment, the avoidance response prevents a shock and terminates the anticipatory warning signal. Shock avoidance and escape from the anticipatory situation are tied together. In the present modification of the classical technique, the subject can behave in either of two alternative fashions: (a) postpone the shock, while at the same time maintaining the warning signal or (b) take the shock while simultaneously terminating the warning signal. The shock can be avoided only at the cost of prolonging the conditions that require the anticipatory avoidance behavior. The anticipatory situation can be terminated only at the cost of receiving a shock.

<sup>1</sup> Now at the Merck Institute for Therapeutic Research.

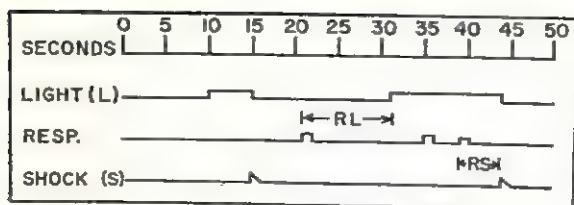


FIG. 1. DIAGRAMMATIC REPRESENTATION OF THE EXPERIMENTAL PROCEDURE (SEE TEXT FOR DETAILS)

The total procedure may, for convenience, be represented symbolically as:  $(RL15'')^D \rightarrow (RS5'')^L, S \rightarrow (RL15'')^D \dots$  This may be read as follows: "In the absence of the light, the response-light interval is 15 seconds. If the animal permits the RL interval to elapse this produces the light, in the presence of which a response shock interval of 5 seconds is in force. Occurrence of a shock terminates the light and reinstates the response-light interval of 15 seconds." Fig. 1 depicts the schedule,  $(RL10'')^D \rightarrow (RS5'')^L, S \rightarrow (RL10'')^D$ . (In the present experiment, the RL interval was 15 sec.)

Each experimental session lasted six to eight hours, with a given animal running no more frequently than once every alternate day. Responses were recorded separately in the presence and in the absence of the light.

### RESULTS

Results for the four Ss are presented in Fig. 2. The control curves in the left-hand frames represent the average response rates during the final sessions of avoidance conditioning, in which the prevailing shock schedule was  $(RS20'')^D$ . The curves in the right-hand frames represent the rate of responding in the light and in the absence of the light following the change in procedure. In three of the four cases the response rate in the stimulus light displayed an initial increase over the rate in  $(RS20'')^D$ , but in all cases the rate gradually dropped to a level well below the control rate. The response rate in the absence of the light did not reveal any marked trend from session to session that was consistent among all the Ss, but, in all cases, ended up higher than the rate during the warning stimulus.

### DISCUSSION

On the basis of previous findings, we might have expected a high rate of avoidance re-

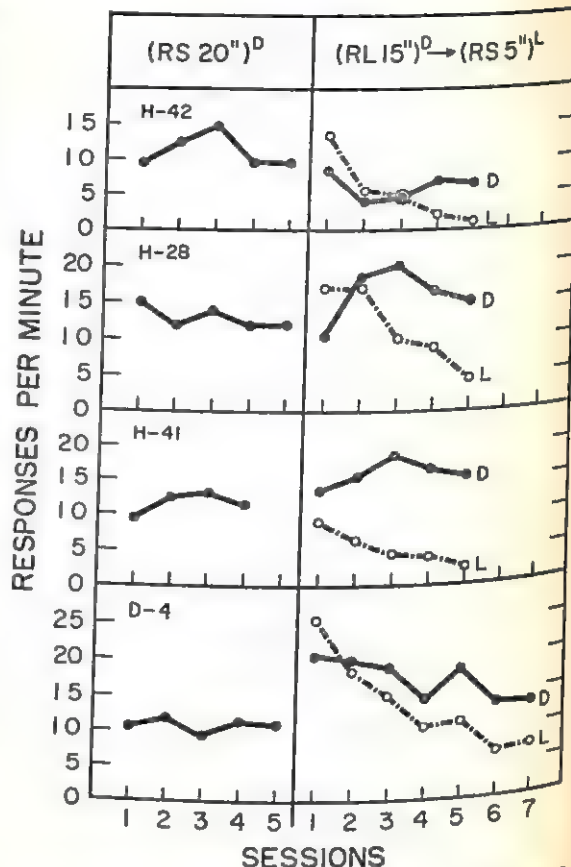


FIG. 2. RATE OF AVOIDANCE RESPONDING IN THE LIGHT (L) AND IN THE ABSENCE OF THE LIGHT (D) FOR 4 ANIMALS

The curves on the left are for the nondiscriminated avoidance procedure, and the curves on the right are for the chaining procedure. Session 1 on the chaining procedure followed immediately after the last session on the nondiscriminated procedure.

sponding in the presence of the warning stimulus and a low rate in its absence. This outcome would have been consonant with the findings of Mowrer and Lamoreaux (8, 9) in the classical avoidance situation and with those of Sidman in a situation in which the avoidance responses could postpone the warning stimulus (14). It has also been shown previously (6, 12, 13) that the rate of avoidance responding is an inverse function of the response-shock interval, within certain limits. On this basis alone, a higher response rate during the warning stimulus would be predicted, when the response-shock interval was only five seconds, than during the control phase, when the response-shock interval was 20 seconds.

In none of the previous experiments, however, was the S faced with the dilemma of



choosing between the warning stimulus and the shock. The response that avoided the shock was also responsible for termination of the warning signal. The present case, in which such a choice was required, produced behavior whose pattern was diametrically opposite to that which is customarily observed. The rate of avoidance responding was low in the presence of the warning stimulus and high in its absence. This was the pattern to be expected if the shock was preferable to the signal. For while lever pressing during the warning stimulus did succeed in postponing the shock, it also served to prolong the stimulus. Behavior other than lever pressing, however, was the only means of terminating the warning stimulus. The low rate of lever pressing indicates that other behavior did, indeed, become prepotent, even though an inevitable shock was one of the consequences of such behavior.

## EXPERIMENT II

If the decline in the rate of lever-pressing during the stimulus was due to the reinforcement of competing behavior through the termination of the stimulus, it should be possible to restore the lever-pressing rate by terminating the light independently of any behavior. This was accomplished by permitting the stimulus light to remain on for five minutes, regardless of whether or not the animal pressed the lever.

### Subjects and Apparatus

The Ss were five male albino rats. The apparatus was the same as in Experiment 1, except for the shock power supply. In this experiment an Applegate shock generator set at maximum output was used. At this setting, the shock varied from 3 to 5 ma.

### Procedure

**Shock terminates stimulus.** This procedure was the same as the "chained avoidance contingencies under stimulus control" procedure described in Experiment I. The salient feature here is the termination of the stimulus coincidentally with the delivery of the shock. All animals were run on the schedule,  $(RL20'')^D \rightarrow (RS10'')^L$ ,  $S \rightarrow (RL20'')^D$ ... for a total of 100 hours. Rat AA-1 was also run on the schedule,  $(RL20'')^D \rightarrow (RS20'')^L$ ,  $S \rightarrow (RL20'')^D$ ... for 100 hours. Rat GF-81 was also run for only one session on the schedule,  $(RL20'')^D \rightarrow (RS5'')^L$ ,  $S \rightarrow (RL20'')^D$ ... Unfortunately, this animal died on the day following this session, but his data were included with the others.

**Five-minute stimulus duration.** On this procedure, whenever the animal permitted the stimulus to appear it remained on for five minutes, regardless of the number

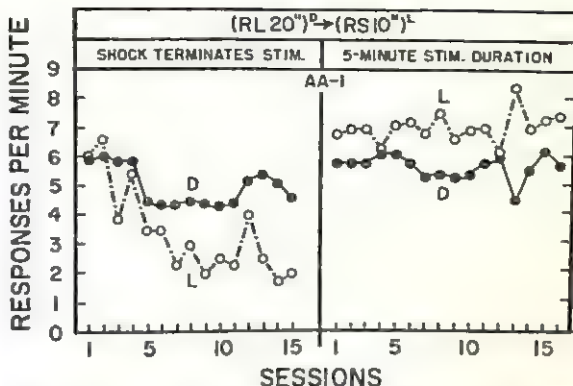


FIG. 3. RATE OF AVOIDANCE RESPONDING IN THE LIGHT (L) AND IN THE ABSENCE OF THE LIGHT (D) FOR RAT AA-1. THE SESSIONS ARE CONSECUTIVE

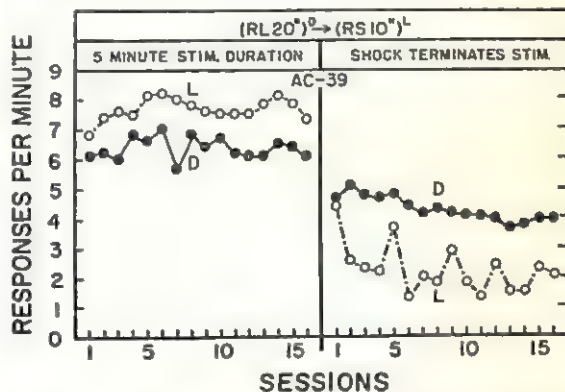


FIG. 4. RATE OF AVOIDANCE RESPONDING IN THE LIGHT (L) AND IN THE ABSENCE OF THE LIGHT (D) FOR RAT AC-39. THE SESSIONS ARE CONSECUTIVE

of responses emitted or the number of shocks delivered. Termination of the stimulus was based upon the passage of time alone, and was independent of the animal's behavior. In terms of our symbolic notation, this procedure might be expressed as  $(RL20'')^D \rightarrow (RS10'')^L$ ,  $(5\text{-min.})^L \rightarrow (RL20'')^D$ ...

The RL and RS intervals for each animal were the same as those described above in the "shock terminates stimulus" procedure. Every animal was run for 100 hours on each schedule. On the procedure involving  $(RL20'')^D \rightarrow (RS10'')^L$ , rats GF-81 and AA-1 ran first on the "shock terminates stimulus" procedure, after which they were changed to the "five-minute stimulus duration" procedure. All the other animals, along with GF-81 and AA-1 on the procedures that involved  $(RS5'')^L$  and  $(RS20'')^L$ , respectively, ran on the "five-minute stimulus duration" procedure first.

## RESULTS

Fig. 3 and Fig. 4 give a session-by-session picture of the response rates under both pro-

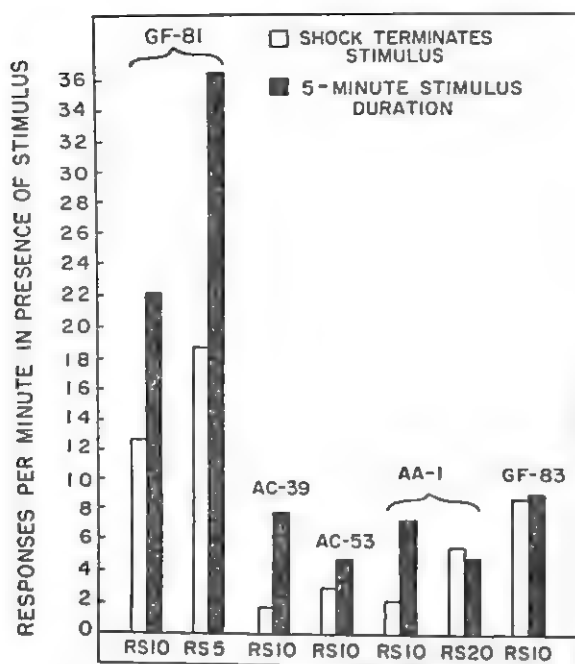


FIG. 5. RATE OF AVOIDANCE RESPONDING IN THE PRESENCE OF THE LIGHT WHEN THE LIGHT TERMINATES WITH THE SHOCK AND WHEN THE DURATION OF THE LIGHT IS INDEPENDENT OF THE SHOCK

The RL interval is 20 sec. in all cases.

cedures for two of the animals. Rat AA-1, which started on the "shock terminates stimulus" procedure, replicated the data presented in Experiment I. The response rate in the presence of the stimulus fell off relative to its own initial level and to the rate in the absence of the stimulus. When the stimulus duration was fixed at five minutes and was, therefore, independent of the animal's behavior, there was a marked increase in the rate of responding during the stimulus. Rat AC-39, started on the "five-minute stimulus duration" procedure, showed essentially the same results. When the stimulus termination was made dependent upon failure to press the lever and receipt of a shock, the response rate in the stimulus dropped markedly. In both animals there was only a slight change in response rate in the absence of the stimulus.

A summary of the data for all animals is presented in Fig. 5. This figure shows the median response rate in the stimulus over the last four sessions of each procedure for each animal. In all cases the RL interval was 20 seconds. The RS intervals employed are indicated beneath the corresponding bars of the

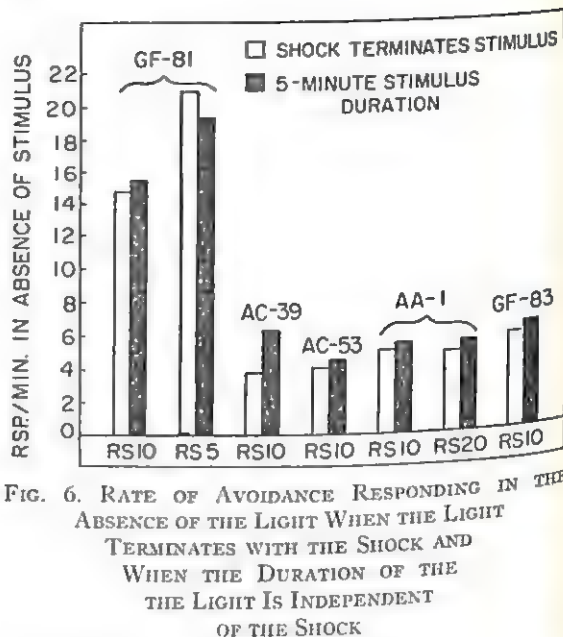


FIG. 6. RATE OF AVOIDANCE RESPONDING IN THE ABSENCE OF THE LIGHT WHEN THE LIGHT TERMINATES WITH THE SHOCK AND WHEN THE DURATION OF THE LIGHT IS INDEPENDENT OF THE SHOCK

The RL interval is 20 sec. in all cases.

graph. Of the five cases in which the RS interval was 10 seconds, the rate during the stimulus was higher when the stimulus duration was independent of the animal's behavior. In the case of GF-83, however, the difference was too small to be taken seriously, and this animal must be considered an exception.

When rat GF-81 was run on an RS interval of five seconds there was again a striking difference in rate between the two procedures. It must be remembered, however, that GF-81 had only one session on (RS5)<sup>2</sup> and that the rates depicted here are undoubtedly not the final levels that would have been reached if additional sessions could have been run. A most interesting finding is that yielded by Rat AA-1 when the RL and RS intervals were both 20 seconds. In this case, it made little difference whether or not the stimulus duration depended upon the S's behavior. Response rates in the presence of the stimulus were essentially the same in both procedures.

The effect of the procedural differences upon the response rate in the absence of the stimulus is shown in Fig. 6. With one exception, the rate on the five-minute stimulus duration procedure was higher, but in no case were the differences for a given animal as great as those displayed by the response rates in the presence of the stimulus depicted in Fig. 5. The one exception, GF-81 at an RS interval of five seconds, probably does not represent a true stable level



of responding because only one session was included.

### DISCUSSION

These data substantiate the view that the decline in lever-pressing rates during the stimulus, observed in Experiments I and II, is due to the reinforcement of nonlever-pressing behavior by the termination of the stimulus; this, in spite of the fact that a shock accompanied stimulus termination. When the stimulus duration was fixed at five minutes, regardless of the animal's behavior, the response rate in the presence of the stimulus increased. In this procedure it was impossible for the animal to "take the shock in order to terminate the stimulus," and the lever-pressing rate assumed a value consistent with the prevailing response-shock interval.

Most current explanations of discriminated avoidance behavior attribute the maintenance of the avoidance response to a "drive" or "aversive" function acquired by the warning stimulus as a result of its temporal relation to the shock (e.g., 5, 7, 10, 17). Termination of the stimulus is held to provide reinforcement for the avoidance response. The present data seem to indicate that the stimulus-shock pairings actually make the warning stimulus more aversive than the shock from which it derived its function. But the data of Rat AA-1 do not support the suggestion that the aversive function of the warning stimulus derives simply from its association with the shock. Rather than dismiss the performance of AA-1 as an example of unexplained variability, we may view the apparent exception with greater than ordinary interest. Rat AA-1 displayed the expected difference in responding under the two procedures when the RL interval was 20 seconds and the RS interval was 10 seconds. There was no large difference, however, when both intervals were equal. In spite of the fact that the warning stimulus was inevitably paired with shock, there is little indication that the stimulus was more aversive than the shock when RL and RS intervals were each 20 seconds. Some factor in addition to the stimulus-shock pairings must be involved in determining the relative aversiveness of stimulus and shock.

Sidman and Boren have suggested (16) that the discriminated avoidance situation may be considered as a "multiple schedule" (1) in which one set of avoidance contingencies pre-

vails in the presence of the warning stimulus and another in its absence. In the present experiments, the two sets of avoidance contingencies may be specified in terms of the RL and RS intervals, the latter prevailing in the presence of the warning stimulus, and the former in its absence. When both intervals were 20 seconds, there was no great advantage to the animal in taking the shock and terminating the stimulus, for the temporal contingencies were the same in the presence and absence of the stimulus. A relatively high rate of lever pressing was therefore maintained in the stimulus, and the shift to a fixed stimulus duration produced no appreciable change. (The fact that this situation obtained when both intervals were numerically equal may be regarded as fortuitous. It is possible that for other animals a much larger difference between the RL and RS intervals would be required to produce a decline in rate during the stimulus.) Other experiments have served to confirm this explanation. It was found that the aversiveness of the warning signal is a function of the RS interval that prevails in the presence of the signal, of the RL interval that prevails in its absence, and of the difference between the two (15). It is necessary to take into account not only the anticipatory situation from which the *S* escapes, but also the conditions that obtain subsequent to the escape. If the latter are relatively favorable, the anticipatory signal may actually become more aversive than the shock itself.

### SUMMARY

An avoidance situation was arranged in which the animals could either (a) postpone the shock and prolong the warning signal or (b) take the shock and terminate the warning signal. With this procedure, we observed a low rate of avoidance responding in the presence of the warning signal and a high rate in its absence. The animals behaved in such a way as to terminate the stimulus as quickly as possible, even though a shock accompanied each stimulus termination. When the signal duration was made independent of the animal's behavior, the rate of avoidance responding in the stimulus increased. It was suggested that the avoidance contingencies in the presence and absence of the warning stimulus are important determiners of the relative aversiveness of stimulus and shock.

## REFERENCES

1. FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement*. Appleton-Century-Crofts, in press.
2. GWINN, G. T. The effects of punishment on acts motivated by fear. *J. exp. Psychol.*, 1949, **39**, 260-269.
3. HILL, H. E., KORNETSKY, C. H., FLANARY, H. G., & WIKLER, A. Effects of anxiety and morphine on discrimination of intensities of painful stimuli. *J. clin. Investig.*, 1952, **31**, 473-480.
4. KAMIN, L. J. The effects of termination of the CS and avoidance of the US on avoidance learning. *J. comp. physiol. Psychol.*, 1956, **49**, 420-424.
5. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
6. MORSE, W. H., & HERRNSTEIN, R. J. The maintenance of avoidance behavior using the removal of a conditioned positive reinforcer as the aversive stimulus. *Amer. Psychologist*, 1956, **11**, 430. (Abstract)
7. MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem-solving." *Harv. educ. Rev.*, 1947, **17**, 102-148.
8. MOWRER, O. H. & LAMOREAUX, R. R. Avoidance conditioning and signal duration: a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, **54**, No. 5 (Whole No. 247).
9. MOWRER, O. H., & LAMOREAUX, R. R. Conditioning and conditionality (discrimination). *Psychol. Rev.*, 1951, **58**, 196-212.
10. SCHOENFIELD, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. H. Hoch & J. Zubin (Eds.), *Anxiety*. New York: Grune & Stratton, 1950. Pp. 70-99.
11. SIDMAN, M. Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science*, 1953, **118**, 157-158.
12. SIDMAN, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *J. comp. physiol. Psychol.*, 1953, **46**, 253-261.
13. SIDMAN, M. On the persistence of avoidance behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 217-220.
14. SIDMAN, M. Some properties of the warning stimulus in avoidance behavior. *J. comp. physiol. Psychol.*, 1955, **48**, 444-450.
15. SIDMAN, M. Conditioned reinforcing and aversive stimuli in an avoidance situation. *Trans. N. Y. Acad. Sci.*, 1957, **19**, 534-544.
16. SIDMAN, M., & BOREN, J. J. A comparison of two types of warning stimulus in an avoidance situation. *J. comp. physiol. Psychol.*, in press.
17. SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.
18. SOLOMON, R. L., KAMIN, L. J., & WYNNE, L. C. Traumatic avoidance learning: the outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, **48**, 291-302.

Received February 14, 1957.



# SELECTIVE RECALL IN SCHIZOPHRENIA AND ITS RELATION TO EGO STRENGTH<sup>1</sup>

ARTHUR S. TAMKIN<sup>2</sup>

*Duke University*

**T**HE PURPOSE of this study is to investigate the nature of selective recall in schizophrenic patients. The hypotheses offered are that schizophrenics manifest a distinctive type of memory process in contrast to normal subjects, resulting in the preferential remembering of experiences connoting personal failure and diminished self-esteem, and that this process may be related to deficient ego strength.

This characteristic memory process may be inferred both from clinical observations of schizophrenics and from theoretical considerations of schizophrenia. Certain aspects of its symptomatology, suggesting selective memory functioning, have been described as brooding preoccupations (4), chronic feelings of insecurity and inadequacy (20), and reaction sensitivity to events which reduce self-esteem (6). Cameron has explained this symptomatology as follows: "In the childhood of both paranoid and schizophrenic patients . . . there have been developed self attitudes which leave the individual unusually reaction sensitive to signs that he may be guilty, inferior, or unworthy in the eyes of others" (5, p. 489).

Although in clinical descriptions and theoretical discussions pathological selective remembering has been designated by different terms, the common implication is that the direction of thinking and remembering in the schizophrenic is determined by his peculiar feelings, needs, and emotional complexes. It seems that he does not passively absorb impinging stimuli and recall his experiences indiscriminately; rather, directive intrapsychic forces predispose him to respond to and retain material which is congruent with his pathological motivations.

In normal functioning, selectivity in both

perception and memory has been demonstrated by a variety of experimental methods, and it has been related to basic needs and defensive mechanisms of the personality. Personal needs and values, for example, have been shown to enhance the perceptual recognition of need- and value-related words (14), and the preferential perception of tasks associated with failure and success has been shown to be determined by individual defensive mechanisms (9, 15). Studies of selectivity in memory are represented by a large body of literature reviewed in several publications (3, 16). The general findings which have emerged are that memory normally functions in a selective manner, enhancing retention of material with which the person is in agreement (7), which is congruent with his value-orientation (8), and which has pleasant affective connotations (12).

A method for measuring selective recall which has enjoyed wide use is the recall of completed and incompleting tasks, introduced by Zeigarnik (22) in 1927. This method is quite suitable for the purposes of this investigation, for when it is applied as a simulated intelligence test in order to enhance the ego involvement of the subjects, the experimenter may systematically manipulate the subjects' experience of self-esteem through success and failure on the tasks. Moreover, motives related to the defense of the ego or other psychologically significant adaptive responses may be evoked. The method of completed and incompleting tasks creates a realistic situation which maintains the subjects' interest by simulating a familiar kind of competitive situation, and is, therefore, a convincing experimental analogue of an emotionally arousing experience which has real personal significance for the participants. Alper (2) has demonstrated that when the completed and incompleting task method is used under ego-involving conditions, individuals who recall a preponderance of successes in this threatening situation are characterized by a "strong ego," whereas those who recall a preponderance of failures have a "weak ego." Several subsequent studies have shown that

<sup>1</sup> This paper is based on a Ph.D. dissertation at Duke University. The author wishes to thank Drs. Louis D. Cohen, George H. Collier, and Gregory Kimble for their invaluable assistance on this project, and Dr. Burke M. Smith of the Roanoke Veterans Administration Hospital who made available patients and facilities.

<sup>2</sup> Now at the Veterans Administration Hospital, Northampton, Mass.

normally the completed tasks are selectively retained (1, 11, 17), and others have confirmed the relationship between this tendency and level of ego strength (9, 10, 13).

When completed and incompleting tasks are presented to schizophrenics, one may hypothesize (a) preferential recall of the incompleting puzzles and (b) a relation between this tendency and deficient ego strength. These specific hypotheses were tested by the experiment reported here.

### METHOD

In order to test the hypothesis that schizophrenics manifest selective recall of failure experiences, their recall of completed and incompleting puzzles which were administered ostensibly as an intelligence test was analyzed. An incidental rather than an intentional (set-to-learn) learning situation was utilized because of its superiority for the demonstration of selective remembering (1). The condition of ego involvement, created by simulated intelligence testing, was established in order to enhance the probability that incompleting puzzles would connote failure, thus presenting a threat to self-esteem. The second hypothesis, which states that the selective recall of schizophrenics is related to deficient ego strength, was explored with the Scott-Duke Questionnaire (18), designed to measure the trait of self-acceptance, a personality characteristic presumably related to level of ego strength. An inverse relation between scores on the Self-acceptance scale of this questionnaire and a preponderant recall of incompleting puzzles would be considered confirmative of the second hypothesis.

### Subjects

An experimental and a control group of 24 Ss each were used in the experiment. The experimental group was composed of male schizophrenics, patients at the Veterans Administration Hospital, Roanoke, Virginia, randomly selected on the basis of availability for psychological testing and ability to cooperate on the testing procedures. Four Ss had to be eliminated because of failure to cooperate. The diagnoses were made by the psychiatric staff. All patients used were recent admissions who had received no psychiatric treatment during their current hospitalization at the time of the experiment. The control group consisted of 24 adult males who had never received psychiatric treatment and who claimed that they did not require such treatment. They were considered to represent a group of normal individuals. They were selected according to

TABLE 1  
VARIABLES ON WHICH THE SCHIZOPHRENIC  
AND CONTROL GROUPS WERE MATCHED

Variable	Schizophrenics		Controls		t test
	N	Mean	N	Mean	
Education	—	8.92	—	9.25	.43
Age	—	30.08	—	29.75	.18 <sup>a</sup>
Negro	4		4		
White	20		20		

Note.—Subjects of both groups resided in the same locality and all were males.

<sup>a</sup> *p* is greater than .20.

criteria of color, age, education, sex, and geographical residence so as to yield a group comparable to the experimental group for these personal variables. The data are presented in Table 1.

The members of the control group were obtained from three separate sources in response to a solicitation to donate one hour of time to an unspecified research project. Eleven Ss volunteered at a labor union meeting, ten volunteered at a state employment office for a monetary incentive, and three were obtained while waiting for relatives at an outpatient clinic.

### Materials

The materials for the completed and incompleting tasks consisted of 10 jigsaw puzzles, each depicting a common animal with its name printed conspicuously along the border. The animals were kangaroo, giraffe, parrot, bear, rooster, eagle, swan, tiger, geese, and elephants.

The Scott-Duke Questionnaire comprises the *L* and *K* scales of the MMPI, which are used as a check on validity, and a Self-acceptance scale. It is a self-administered, true-false type of questionnaire, scored by means of an objective key. According to Scott (18), the Self-acceptance scale was based upon Sheerer's (19) definition of the self-accepting person, and it was shown to have a high reliability and validity as a measure of self-acceptance. The split-half reliability was reported by Scott as .94. Its validity was predicated upon its correlation of .77 (.83, corrected for attenuation) with Rogers' Q-sort technique with 58 Ss. The empirical finding that it differentiated between a variety of psychiatric disorders ranging from the psychoneuroses to schizophrenia, which yields the lowest score, as well as its relation to the Q-sort measure of self-acceptance, tends to justify its use in this experiment as an estimate of ego strength.

### Experimental Procedure

All Ss in both the experimental and control groups underwent an identical procedure, but only 17 Ss from each group, matched for the personal variables, received the Scott-Duke Questionnaire. The procedure consisted of the following sequence of individually administered tests: a vocabulary test, assembly of jigsaw puzzles, digit span test, test for recall of jigsaw puzzles, and the Scott-Duke Questionnaire. The vocabulary and digit span tests were introduced in order to simulate more plausibly an intelligence testing condition, and



the instructions given to both groups were worded appropriately. The ten jigsaw puzzles, constituting the memory material, were administered in a random sequence. The order of completions (C) and incompletions (I) was identical for each *S* in the sequence ICCIICIC. The *Ss*, timed for their performance on each puzzle, were allowed to complete every one designated as a C puzzle, and were prevented from completing every I puzzle by the imposing of an arbitrary time limit just before the last two pieces were fitted into the partially assembled puzzle. After completing each C puzzle, the *S* was asked to tell what the picture was if he did not do so spontaneously. In order to ensure equal exposure time to all puzzles, the I puzzles, which were finally completed by the experimenter, were exposed to the *S*'s view, and he was asked to state what the picture was. Immediately after the interpolated digit span test, each *S* was asked to recall the names of all the jig-saw puzzles which he could remember.

### RESULTS

The chi-square test was used to answer the question whether any distinctive pattern of recall was associated with the schizophrenics as compared with the control *Ss*. The *Ss* of each group were divided into three categories according to their recall patterns: (a) the recall of a preponderance of I puzzles (I), (b) recall of a preponderance of C puzzles (C), and (c) recall of an equal number of I and C puzzles (Equal). The first two patterns represent different types of selective recall, while the third pattern shows no selectivity. The distribution of *Ss* falling into each category is shown in Table 2. Chi square for the association of recall pattern with psychiatric status (normal or schizophrenic) was 6.89 ( $.05 > p > .02$ ). It may be observed that while a majority of the schizophrenic group demonstrated the I pattern, a minority of the control group also did so. The majority of the control group was distributed evenly between the Equal and C patterns.

In order to determine the relationship between scores on the Self-acceptance scale and the three recall patterns, all *Ss* were sorted according to their recall patterns and an analysis of variance of their Self-acceptance scores performed. The obtained *F* of .32, as shown in Table 3, is not significant.

Since Self-acceptance scores were unrelated to the recall patterns, evaluations were made of the reliability and applicability of the scale in the assessment of ego strength with the *Ss* of this experiment. Internal consistency was estimated by the split-half method, using the

TABLE 2  
FREQUENCIES OF SUBJECTS DEMONSTRATING  
EACH TYPE OF RECALL PATTERN

	I	Equal	C
Schizophrenics	15	4	5
Controls	6	9	9

TABLE 3  
SUMMARY OF ANALYSIS OF VARIANCE OF  
SELF-ACCEPTANCE SCORES

Source of Variance	df	Sum of Squares	Mean sum of Squares	<i>F</i>	<i>P</i>
Between groups	2	31.68	15.84	.32	> .05
Within groups	31	1544.08	49.81		
Total	33	1575.76			

Spearman-Brown formula. A coefficient of .90 was obtained for the 34 *Ss* in the combined groups. This coefficient, similar to that of .94 reported by Scott, is high enough to suggest that the scale has high internal consistency. While a more meaningful evaluation of reliability would involve the question of the stability of the measurements over time, test-retest data were not available to permit this type of evaluation. With regard to the applicability of the Self-acceptance scale in estimating ego strength, on a priori grounds it should differentiate between hospitalized schizophrenics and normal persons. The mean Self-acceptance score of the control group was 30.65, and of the schizophrenic group 25.12. A *t* of 2.96 was significant at the .01 level, but since there was considerable overlap between the two distributions, its effectiveness in discrimination was rather low.

### DISCUSSION

While the hypothesis that schizophrenics selectively recall experiences connoting personal failure and diminished self-esteem was supported by the experimental findings, the hypothesis of the relation of this memory process to deficient ego strength was unsupported. The failure to confirm the second hypothesis may be attributable, in part, to two possible sources of error: one having to do with deficiencies in the instrumentation employed, the other involving certain weaknesses of the concept of ego strength.

The estimation of ego strength by means of

the Self-acceptance scale seemed tenable in view of its demonstrated relationships to psychopathology and to an independent, objective measure of self-acceptance. However, its power of discrimination proved to be quite low owing to the degree of overlap of scores in the schizophrenic and normal groups. Furthermore, while high split-half reliability was demonstrated, the question of the stability of measurement over time was not evaluated owing to the unavailability of test-retest data. This would seem to be the more important consideration in establishing the level of reliability because of the probable susceptibility of the schizophrenics to transient influences and momentary moods which would make their responses to the Self-acceptance scale undependable. In view of these sources of experimental error, one may question the effectiveness of the Self-acceptance scale as an estimate of ego strength.

Another factor which might be responsible for the finding of no relationship, in spite of the positive findings of other experimenters who used different instruments and normal Ss, is that due to ambiguous operational and conceptual definitions of ego strength, different facets of this trait are measured by the various instruments employed. The findings of a recent study comparing several independent measures of ego strength have drawn attention to this ambiguity on empirical grounds (21). Because ego strength, defined by Erikson as "the individual's capacity for appraising the reasonable limits in his interpretations and perceptions of his environment" (10, p. 46) and measured by Rorschach  $F+$  %, has been shown to correlate significantly with selective recall, this is no indication that ego strength measured by the Self-acceptance scale should. This consideration argues the need for a more precise definition of the presently vague concept of "ego strength." Certainly, the negative finding obtained in this experiment points up the need for continued research on the problems of ego strength and its role in schizophrenia.

#### SUMMARY

Two hypotheses concerning certain aspects of schizophrenic symptomatology were tested: (a) Schizophrenics manifest a distinctive type of memory process which results in the selective recall of experiences connoting personal failure

and diminished self-esteem and (b) this process is related to deficient ego strength. The method used to test the first hypothesis was to administer jigsaw puzzles as a simulated intelligence test, allowing each S to complete only half of the group of puzzles, and to determine whether there occurred selective recall by the schizophrenics for the incompleting puzzles connoting personal failure. As a test for the second hypothesis the Scott-Duke Questionnaire, which contained a Self-acceptance scale and two validity scales from the MMPI, was administered to provide a measure of ego strength, and its power of discrimination between Ss grouped according to their recall pattern was ascertained. The Ss consisted of 24 schizophrenic patients who were recently admitted or readmitted for hospitalization and a control group of 24 normal adults matched for age, education, sex, color, and geographical residence. While the first hypothesis was confirmed at a significant level of confidence, the second was not supported by the experimental data. The failure to confirm the second hypothesis seemed to be attributable, in part, to two sources of error; one arising from deficiencies in the instrumentation, and the other from certain weaknesses of the concept of ego strength. It is conceivable that due to the vague operational and conceptual definitions of ego strength, different facets of this trait may be measured by the various instruments which have been employed. It is believed that the negative finding concerning the second hypothesis points up the need for continued research on the problem of ego strength and its role in schizophrenia.

#### REFERENCES

1. ABORN, M. The influence of experimentally induced failure on the retention of material acquired through set and incidental learning. *J. exp. Psychol.*, 1953, **45**, 225-231.
2. ALPER, THELMA G. Memory for completed and incompleting tasks as a function of personality: an analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403-420.
3. BARRETT, DOROTHY M. Memory in relation to hedonic tone. *Arch. Psychol.*, 1938, No. 223.
4. BLEULER, E. *Dementia praecox or the group of schizophrenias*. (Translated by Joseph Zinkin.) New York: International Universities Press, 1950.
5. CAMERON, N. *The psychology of behavior disorders*. Boston: Houghton Mifflin, 1947.
6. CAMERON, N., & MAGARET, ANN. *Behavior pathology*. Boston: Houghton Mifflin, 1951.



7. CLARK, K. B. Some factors influencing the remembering of prose material. *Arch. Psychol.*, 1940, No. 253.
8. COFER, C., & HAVRON, D. Influence of personal values on learning. *Amer. Psychol.*, 1950, 5, 274. (Abstract)
9. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, 47, 230-235.
10. ERIKSEN, C. W. Psychological defenses and "ego strength" in the recall of complete and incompleting tasks. *J. abnorm. soc. Psychol.*, 1954, 49, 45-50.
11. KENDLER, TRACY S. The effect of success and failure on the recall of tasks. *J. gen. Psychol.*, 1949, 41, 79-87.
12. LAFFAL, J. The learning and retention of words with association disturbances. *J. abnorm. soc. Psychol.*, 1952, 47, 454-462.
13. LAZARUS, R. S., & LONGO, N. The consistency of psychological defenses against threat. *J. abnorm. soc. Psychol.*, 1953, 48, 495-499.
14. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, 43, 142-154.
15. POSTMAN, L., & SOLOMON, R. L. Perceptual sensitivity to completed and incompleting tasks. *J. Pers.*, 1950, 18, 347-357.
16. RAPAPORT, D. *Emotions and memory*. New York: International Universities Press, 1950.
17. RUSSELL, W. A. Retention of verbal material as a function of motivating instructions and experimentally induced failure. *J. exp. Psychol.*, 1952, 43, 207-216.
18. SCOTT, H. A. The empirical assessment of self-acceptance. Unpublished doctoral dissertation, Duke Univer., 1953.
19. SHEERER, ELIZABETH. An analysis of the relation between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, 13, 169-175.
20. SULLIVAN, H. S. Research in schizophrenia. *Amer. J. Psychiat.*, 1929, 9, 553-567.
21. TAMKIN, A. S. An evaluation of the construct validity of Barron's ego-strength scale. *J. clin. Psychol.*, 1957, 13, 156-158.
22. ZEIGARNIK, B. Das Behalten erledigter und unerledigter Handlungen. *Psychol. Forsch.*, 1927, 9, 1-85.

Received February 4, 1957.

# EFFECTS OF PERCEIVED DEPENDENCY RELATIONSHIPS UPON CONFORMITY TO GROUP EXPECTATIONS<sup>1</sup>

LEONARD BERKOWITZ

*University of Wisconsin*

TWO PREVIOUS papers have shown that (a) groups evaluated as a unit become more highly motivated than groups whose members are evaluated separately (1), and (b) the individually evaluated groups that develop the highest task motivation are those whose members initially are highly motivated and then evaluated unfavorably (2). In both cases it was suggested that the final high task motivation resulted from a shared perception of group member interdependence for reward attainment. The present experiment provides a direct test of this hypothesis.<sup>2</sup> In addition, an attempt will be made to demonstrate that the hypothesized relationship is a special case of a

more general relationship between perceived member interdependence and motivation to conform to relevant group expectations.

Several components of perceived interdependence must be considered in accounting for its effect upon group-task motivation. The writer has proposed elsewhere (2) that the degree to which an individual feels dependent upon others is a joint function of two factors: the judged probability that his progress toward his goal is dependent upon the activity of these others (perceived instrumentality) and the extent to which he values this goal.<sup>3</sup> In all likelihood, perceived instrumentality and goal value affect the degree of perceived dependence regardless of the particular direction of the dependency relationship. Thus, assuming that each group member's task performance is necessary for any single member's goal attainment, a subject (S) should see himself as being highly dependent upon his teammates when he is likely to gain a goal he values highly through group activity with them. He also should perceive his teammates as being highly dependent upon him when the group activity gives them access to a goal they would value highly. Any given member of an *interdependent* group probably has both of these perceptions; namely, that he is dependent upon the others for the attainment of his goals, and also that they are dependent upon him for their goal attainment.

The primary motive for an individual who can attain his task-associated goal independently of the others in the group (i.e., an S who perceives low instrumental interrelationships among the group members) stems solely from the value of this goal to him. The member of an instrumentally interrelated group, however, has additional motives created by the

<sup>1</sup> This study was supported by funds granted by the University Research Committee of the University of Wisconsin. The author would like to acknowledge Mr. Sidney Willmuth's invaluable assistance in the collection of the data.

<sup>2</sup> There is some evidence providing tentative support for the hypothesis. Deutsch (3) has shown, for example, that cooperative groups are more highly task motivated than competitive groups. The individuals in the cooperative groups are described as being "promotively interdependent" in that any member's instrumental behavior could advance all members toward the goal. Deutsch's results would appear to indicate, then, that these interdependent Ss were more highly motivated toward the assigned group task than the less promotively interdependent members of the competitive groups. However, the obtained differences may not be due to the degree of perceived promotive interdependence per se. There is a possibility that the high group-task motivation was produced by one or more of the by-products of the cooperative relationship, such as the high group cohesiveness, either directly or in interaction with the perceptions of member interdependence.

A somewhat similar difficulty is presented by a second related study. Pepitone (8) found that Ss who had been told their jobs were important to the group were more highly task motivated than Ss who had been told their jobs were of relatively little importance. Since group members generally are dependent upon the behavior of those with important jobs, the Ss' task motivation apparently was directly affected by their belief that others in the group were dependent upon them. However, there also are alternative possibilities. It is conceivable that ego-enhancing motives were aroused by E's description of S's job as being important to the group, and that these motives affected Pepitone's results.

<sup>3</sup> Helen Peak has made a similar analysis of attitudes (7). She has suggested that the position any person takes on an attitude scale is some function of: (a) "The judged probability that the [attitude] object leads to good or bad consequences [i.e., the perceived instrumental relationship between the object and the perceiver's goal], and (b) the intensity of the effect expected from those consequences." P. 154.



knowledge that others are dependent upon his task behavior for their rewards.

This knowledge may affect the person's task motivation in at least two different ways. First, if the others are in a position to become aware of his performance, he may feel that effective task behavior on his part is necessary to win their good will. Although the desire to win the approval of others probably is an effective motivating force in itself, this desire should be particularly important when the person is dependent upon the others' task activity for his access to a valued goal, as is the case in interdependent groups. Second, his effective performance may be necessary in order to conform to his interiorized standards of proper social conduct. Since the group members in the experiment to be described observed each other's behavior, it is not possible to test these alternatives unequivocally. However, whatever the reason, under the present conditions the task motivation of Ss should be affected by the perception that others in their group are dependent upon them.

Several specific hypotheses can be formulated on the basis of the foregoing discussion. The first has to do with the perceived instrumental interrelationships among the group members. 1. *Members of groups who are instrumentally interrelated for reward attainment become more highly motivated toward the group task than Ss who can achieve their goals independently of the others in the group.* The remaining hypotheses are based upon the assumption that the value of the goal to S and/or his teammates also affects the perceived dependency relationships. In highly interdependent groups the task-associated goal is likely to be valued highly by all group members. 2. *Under conditions of instrumental interdependence, Ss who are told that both they and their teammates can attain a valued goal through the group activity become more highly task motivated than Ss who are told that either only S or only S's partner can gain this goal.* At least part of S's task motivation in an interdependent group stems from his knowledge that others are dependent upon him for their attainment of a goal they would value highly. 3. *Given that the instrumentally interrelated group members are aware of each other's behavior, Ss instructed that only their partners have access to a valued goal through the group activity become*

*more highly task motivated than Ss working for a goal that is less valued by both themselves and their partners.*

## METHOD

### *Group Task and Measure of Task Motivation*

Two Ss, both male or female, were scheduled together, ostensibly to work on a task assessing their "effective intelligence." They were told that the object of the task was for each of them to transmit a list of nonsense syllables to the other as quickly as possible by means of a simulated telegraph system. Each S operated two telegraph keys controlling light signals on his partner's desk, "dot" for one light, and "dash" for the other. Employing a telegraphic code supplied to each S, S-I sent the first nonsense syllable on his list to his partner, S-II, who decoded and copied the message. S-II then sent the first syllable on his list to S-I, who in turn decoded and wrote out the partner's message before sending his second syllable to S-II. This alternation was continued for 15 syllables by each S. The group members obviously were instrumentally interrelated because of the structure of the task.

After completing a given nonsense syllable, each S flashed a third light on his partner's desk. A trial is defined as the time in seconds taken by a given S to receive and decode his partner's message and then send his own message. The 15 trials were divided into three blocks of five trials each for computational convenience. The E told the Ss that the first five trials were "for practice" and would not count toward their score. The total time for these trials in the first block provided for each S a baseline activity rate from which the time scores for the two remaining blocks were subtracted to yield two improvement scores. The latter scores, the dependent variable in this study, serve as an index of S's motivation to work rapidly in the given block, i.e., his motivation to achieve the designated task objective. Analyses of the data disclose no significant relationship between an S's improvement scores and those of his partner.

### *Research Design and Experimental Manipulations*

Under the pretext that he was being interviewed for life history data, each S was individually given the task instructions appropriate to the condition to which he was assigned. Five conditions were established as follows:

- Both S and his partner eligible for a prize, high instrumental interrelationship.
- S alone eligible for a prize, high instrumental interrelationship.
- S's partner alone eligible for a prize, high instrumental interrelationship.
- No prizes to be attained, high instrumental interrelationship.
- Both eligible, low instrumental interrelationship.

In each group, S-I (randomly chosen) was informed that ten-dollar prizes were to be awarded to the best Ss. Accuracy of transmission was listed as one of the criteria, but speed was emphasized. For Condition 1,

*S-I* was told both he and his partner were eligible for these prizes; the prizes were to be awarded to the best teams. In the second and third conditions, *S-I* was informed that *E* was interested in studying the behavior of only one of the two in the group, and either that he, *S-I*, was the only one being studied and hence eligible (Condition 2) or that his partner was being studied and only he was eligible (Condition 3). *S-II* was never told about the prizes and was instructed only to "try to complete the task quickly." Condition 4 is comprised of these latter *Ss*.

The *Ss*' eligibility for the prize in these conditions of high instrumental interrelationships is assumed to result in a perceived high degree of dependency upon his partner, while the partner's eligibility presumably causes *S* to view his partner as being highly dependent upon him. The first four conditions therefore can be considered as representing a  $2 \times 2$  factorial design, varying the relations of presumed perceived dependence between *S* and his partner for the prize reward.

In the fifth condition, *S-I* was told that he and his partner would be judged independently and that each was eligible for separate prizes. These instructions were designed to lower the perceived instrumental interrelationships between the group members by counteracting the manifest task structure. A questionnaire, consisting of four seven-point rating scales, was administered at the end of the experimental session to test the success of this manipulation in producing the belief that *Ss*'

attainment of the reward was not dependent upon the activity of his partner. A sample item was: "How dependent were you upon your partner in completing your task as quickly as possible?" As is indicated in Table 1 below, the items refer to the perceived *instrumental interrelationships* between the team members, and not to the perceived *degree* of dependence resulting from the interaction of this instrumental interrelationship with goal value. Since the four items were significantly and positively intercorrelated, *Ss*' perceived instrumentality score was determined by summing his four ratings.

All *Ss* were volunteers from liberal arts courses at the University of Wisconsin. There were nine *Ss* in each of the conditions. Three *S-I*s were chosen at random from each of the first three conditions and the *S-II*s paired with them were used for the Condition 4 data. Approximately half of the *Ss* were females distributed evenly among the conditions.

## RESULTS

### *Dependency Differences Created by Differences in Goal Value*

The *Ss* in the first four conditions were led to believe that they and their partners were instrumentally interrelated in completing the assigned task. As is shown in Table 1, Conditions 1 through 4 had essentially similar means on the postsession questionnaire assessing the perceived instrumental relations between the team members. With perceived instrumentality constant, then, the direction and degree of dependency between team members in these conditions presumably varied directly with eligibility for the prize. Analyses of variance of the two sets of improvement scores in Conditions 1 through 4 revealed significant main effects for both *S Eligible for Prize* ( $F = 15.35$ ;  $p < .01$ ), and *S's Partner Eligible for Prize* ( $F = 10.14$ ;  $p < .01$ ). The first main effect attests to the success of the goal value manipulations. The more *S* stands to gain by completing the task rapidly, the faster he works. The second effect suggests that *Ss*' group-task motivation increases when he perceives that other group members are highly dependent upon him for the attainment of their valued goals.

There were no significant differences in time scores for the baseline period among Conditions 1 through 4, so Table 2 presents only the mean improvement scores for the second and third trial blocks. With the exception of the differences between Condition 4 and the remaining conditions in both trial blocks, the effects of the perceived dependency relationships did not

TABLE 1  
PERCEIVED INSTRUMENTAL INTERDEPENDENCY<sup>a</sup>

Condition			
Eligible for Prize	Instrumental Interdependency	Mean	SD
1. Both	High	18.89	6.30
2. <i>S</i> alone	High	20.56	5.23
3. <i>S's</i> partner alone	High	19.89	4.88
4. No prize	High	20.33	5.66
5. Both	Low	14.00	6.06

<sup>a</sup>  $N = 9$  in each condition. There are no significant differences among condition means.

TABLE 2  
MEAN TIME IMPROVEMENT SCORES IN CONDITIONS VARYING IN DIRECTION AND DEGREE OF DEPENDENCY BECAUSE OF DIFFERENCES IN PRIZE ATTAINABILITY

Trials	Prize Attainable By <i>S</i>		Prize Not Attainable By <i>S</i>	
	Prize Attainable By <i>S's</i> Partner (Condition 1)	Prize Not Attainable By <i>S's</i> Partner (Condition 2)	Prize Attainable By <i>S's</i> Partner (Condition 3)	Prize Not Attainable By <i>S's</i> Partner (Condition 4)
Block 2	58.33	57.55	59.00	21.67
Block 3	73.44	55.67	45.78	28.33
All trials	65.89	56.61	52.39	25.00



become significant until Block 3. According to Hypothesis 2, Ss in Condition 1, who were told that both they and their partners were eligible for the valued rewards and, therefore, who may well have perceived a relatively high degree of interdependency, should be more highly motivated than the Ss in the other instrumentally interrelated conditions. This hypothesis was confirmed at an acceptable level of confidence ( $p < .05$  by  $t$  tests) for each of the three comparisons with Condition 1 in the Block 3 data.

The only other significant difference in Block 3 was that between Conditions 3 and 4 ( $p = .05$ ). In support of Hypothesis 3, Ss who were informed that the prize was attainable by their partners became highly motivated to perform their task even though they themselves were not eligible for the reward.

#### *Effects of Perceived Instrumental Interdependency*

Both S-I and his partner were eligible for the prize in Conditions 1 and 5, but the latter condition, it will be recalled, was designed to have low perceived instrumental interrelationships between the team members. Analysis of the perceived instrumentality scores indicated that the low instrumental dependency manipulation was not completely successful; as is shown in Table 1, the Condition 5 mean was not significantly lower than the Condition 1 mean ( $p = .15$ ). A preliminary assessment of the motivational effects of this perceived instrumental interrelationship was obtained by discarding the two Ss with the lowest perceived instrumentality scores from Condition 1 and the two Ss with the highest scores from Condition 5 so that the conditions did differ significantly on this variable by  $t$  test ( $p < .05$ ). The Block 3 difference in task motivation scores became significant with the elimination of these four Ss ( $p < .05$ ), providing tentative support for Hypothesis 1. The improvement score means in Conditions 1 and 5 were now 81.00 and 49.57 respectively. Thus, Ss who believe they and their partners are instrumentally dependent upon each other for the attainment of a valued goal are more highly motivated than Ss who believe they and their partners can attain this goal independently of each other.

#### DISCUSSION

The obtained results generally support the hypotheses of the study. Perceived group member interdependency for reward attainment can increase an individual's motivation toward the assigned group task. The findings also suggest that this heightened motivation can be a cumulative effect reflecting the self-perception that he is dependent upon his partner and also the perception that the partner is dependent upon him.

If we assume that an individual who sees that the others in his group are dependent upon him and who also perceives that these others demand or expect certain task-relevant behaviors of him (in the present case, rapid work), the following general proposition can be formulated: *the strength of the motivation to conform to the task-relevant expectations of others in the group is a positive function of the perceived degree of interdependence among the members.* Intragroup opinion conformity under certain conditions, as well as many instances of behavioral conformity to group norms, can then be seen as special cases of this general proposition. For example, it is clear that pressures toward uniformity of opinion are heightened when "group locomotion" requires unanimity (6). To illustrate, Deutsch and Gerard (4) recently created this type of requirement in some of their experimental conditions by offering prizes to the best groups. The Ss given this instruction probably realized that any group member's chances of obtaining the desired goal was dependent upon the activity (i.e., expressed opinions) of the others in the group. Different instructions produced low perceived interdependency in other groups. As a result, the Ss in the highly interdependent groups exhibited a greater tendency to conform to the opinions of others in the group than the Ss in less interdependent groups even though it was fairly obvious that the group opinion was wrong and they could express their own opinions anonymously.

Thus, perceived high interdependence apparently produces conformity to group opinions as well as high group task motivation, supporting the hypothesis of a general relationship between perceived task interdependency and conformity to the task expectations of the others in the group. Attitudinal variables that

affect the degree to which interdependent members adopt each other's opinions, then, may generally affect behavioral conformity to relevant shared expectations in interdependent groups. Dittes and Kelley (5), employing interdependent groups faced with the requirement of consensus, found that the members' adoption of the group opinion was a function of the extent to which they felt accepted by the group. If the present general proposition is correct, differences in acceptance by the group also should produce differences in the relationship between perceived interdependence and group task motivation.

#### SUMMARY

The present study provides a direct test of an hypothesis invoked to account for the results of two preceding experiments, viz.: group members perceiving themselves as interdependent for their reward attainment become highly motivated toward the group task. An analysis also is made of the component relationships involved in perceived interdependence between members of a two-person group. The results generally are in accord with the following predictions: 1. Members of groups who are instrumentally interdependent become more highly motivated toward the group task than Ss who can achieve their goals independently of the others in the group. 2. Under conditions of instrumental interdependence, Ss who are told that both they and their teammates can attain a valued goal through the group activity become more

highly task motivated than Ss who are told either that only S or only S's partner can attain this goal. 3. Given that the group members are instrumentally interdependent and also in a position to see each other's behavior, Ss told that only their partners have access to a valued goal become more highly motivated toward the task than Ss working for a goal that is less highly valued by both themselves and their partners.

#### REFERENCES

1. BERKOWITZ, L., & LEVY, B. I. Pride in group performance and group-task motivation. *J. abnorm. soc. Psychol.* 1956, **53**, 300-306.
2. BERKOWITZ, L., LEVY, B. I., & HARVEY, A. R. Effects of performance evaluations on group integration and motivation. *Hum. Relat.*, in press.
3. DEUTSCH, M. The effects of cooperation and competition upon group process. *Hum. Relat.* 1949, **2**, 199-231.
4. DEUTSCH, M., & GERARD, H. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629-636.
5. DITTES, J., & KELLEY, H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100-107.
6. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271-282.
7. PEAK, HELEN. Attitude and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1955. Lincoln, Neb.: Univer. of Nebraska Press, 1955. Pp. 148-188.
8. PEPITONE, EMMY A. *Responsibility to the group and its effects on the performance of members*. Unpublished doctoral dissertation, Univer. of Michigan, 1952.

Received February 3, 1957.



# THE EFFECT OF TRAINING UPON ACCURACY AND VARIABILITY IN INTERPERSONAL PERCEPTION<sup>1</sup>

W. J. CROW<sup>2</sup>

*Behavior Research Laboratory, University of Colorado*

ONE purpose of many training programs in clinical, counseling, and personnel psychology is to increase the accuracy of the trainee's perceptions of others. The ability to "understand others," or to have "clinical intuition" is considered a basic requirement for the good practitioner. It has been difficult to evaluate the success of such training, however, because until recently there has been no way to measure, by objective tests, the subtle and complex psychological variables that are presumably detected by the practitioner. Therefore, determination of improvement in clinical judgment has itself depended upon other clinical judgments. That is, either the accuracy of the students' perceptions would be directly rated by observers, or the students' judgments would be compared with the judgments of experienced practitioners.

Recent methodological improvements in the measurement of interpersonal perceptiveness have made it possible to measure the accuracy of clinical judgment in a more objective way. It is now customary to ask a person (the object) to fill out a personality questionnaire and then ask the trainee (the subject) to fill out the identical questionnaire just as he believed the object has filled it out. The measure of the subject's accuracy is the degree of correspondence between the object's actual responses and those which the subject estimated that the object had made. Such a technique has been used by Kelly and Fiske (7), for example, to measure the diagnostic competence of clinical trainees.

Early studies using this technique for measuring interpersonal perception focused primarily upon accuracy, but attention has been called recently to the importance of other components of the interpersonal-per-

ception situation (1, 2, 5). Reasoning mathematically from an analysis-of-variance model, Cronbach (1) has advanced a hypothesis for the relationship between accuracy and one of these components—the amount of variability in the subjects' estimations:

There is an optimal degree of differentiation in making judgments. If a judge [subject] can make accurate judgments as to the relative location of others [objects] on a continuum, then he is wise to make  $\sigma_y$  (the variability in his estimations) as large as  $\sigma_x$  (the variability in the others' scores)—never larger. But if he is forced to base his judgment on inadequate cues or if the available personality theory and situational knowledge do not permit trustworthy inference, then he should treat people as if they were very nearly alike. The person who attempts to differentiate individuals on inadequate data introduces error even when the inferences have validity greater than chance (1, p. 181).

It follows from Cronbach's hypothesis that, if the subject increases the variability of his judgments from one time to another and his ability to make accurate judgments does not increase correspondingly, then he will make greater errors the second time than the first. It also follows that whenever a judgmental task is very difficult or the subjects are very low in ability, subjects with little variability in their estimations will make small errors, and subjects with high variability will make large errors. Stated otherwise, accuracy and variability will be negatively correlated when the task is difficult or the subject is very inaccurate.

The purpose of the study herein reported was twofold: (a) to investigate the effects of a training program upon interpersonal perception, and (b) to test the hypothesis of the relationship between accuracy and variability in estimations.

To accomplish these purposes, interpersonal perception measures were administered at the beginning, during, and after the Senior year to a class of medical students who were divided into experimental and control groups. The experimental group was instructed in physician-patient relationships and was provided with the opportunity to establish such relationships through more prolonged contact

<sup>1</sup> This investigation was carried out as part of a larger study of medical education supported by the Commonwealth Fund of New York. The author is indebted to Dr. Kenneth R. Hammond for assistance and advice and to Dr. Fred Kern, Jr., for his assistance during many stages of the study.

<sup>2</sup> Now at San Diego College.

with patients than was possible for the control group. It was expected (a) that the students in the experimental group would become more accurate in interpersonal perception than students in the control group, and (b) that variability in estimation would be negatively related to accuracy of judgment.

## METHOD

### *Subjects and Treatments*

The Ss were 72 Senior medical students who were assigned at random to experimental and control groups. Approximately one fourth of the students were assigned to the experimental clinic during the first half of the year, and the remaining three-fourths were assigned to one of two surgery clinics or to an internal-medicine clinic. During the second half of the year, one of the first-half surgery groups was assigned to the experimental clinic, the other was assigned to the internal-medicine clinic, and the first-half experimental and internal-medicine groups were assigned to the surgery clinics. All students were tested at the beginning (Pretest), at the midpoint (Posttest I), and at the end (Posttest II) of their Senior year.

### *Measures*

Thirty-six patients were selected at random<sup>3</sup> (with certain restrictions) from the medical wards of two hospitals, and each was interviewed by a physician. Six-minute sound films were taken of each of these interviews, and three randomly-selected sets of ten films were shown to the Ss, one of them at each of the three test administrations.

The Ss were required to perform two tasks: Task A was to estimate the self-ratings of the patients on seven personality scales, which were paraphrases of selected Minnesota Multiphasic Inventory scales (6); Task B was to estimate where the patient "really was" on the same set of scales. The patient's self-ratings were the criteria of accuracy on Task A; the patients' performance on the MMPI for Task B.<sup>4</sup>

### *Scoring*

The Random-Comparison Method (2) was used to compute the scores, since this method yields a Differential-Accuracy score by eliminating artifactual components from the raw scores. In addition to the Accuracy scores, two response-set measures, which had previously been found to be stable aspects of the students' performance (4), were investigated in the present study: 1. *A measure of adherence to stereotype*, defined as the inverse of the Average-Variance score of an S's estimations for the scales. (An S may perceive wide differences among people in the characteristic under consideration, or he may perceive people as being very

much alike. He may, in other words, cluster his estimations closely or he may vary them widely. If an S clusters his estimations closely, his Average-Variance score will be small and his adherence to stereotype will be high. The variance of the estimations for each scale was computed and averaged for all scales as a measure of perceived individual differences.) 2. *A measure of assumed veridicality*, defined as the inverse of the Average Difference between the S's estimation for the object's self-rating and the S's estimation for where the object "really was." (An S could assume a large measure of agreement between an object's self-rating and "reality," or he could assume little agreement. If an S assumed a large measure of agreement, the Average-Difference score between his estimation for the object's self-rating and his estimation for where the object "really was" would be small, and the assumed veridicality would be high.)

The Ss' accuracy, adherence to stereotype, and assumed veridicality were measured in this fashion for each of the three administrations. Change scores were computed from pre- to posttest in order to compare the experimental and control groups and to permit investigation of the relationship between accuracy and the two response-set characteristics.

### *Effect of Training upon Interpersonal Perception*

Table 1 presents the means and standard deviations of the change in the Ss' performance between the Pretest and Posttest I (first-half change) and between the Posttest I and Posttest II (second-half change) for the Accuracy, Average-Variance, and Average-Difference scores. The significance of the difference between the mean change for the experimental and control groups was computed by use of the *t* test for small samples.

Examination of the accuracy results revealed that the control-group improvement was significantly greater than the experimental-group improvement for one of the four measures (Task B, first half). The control-group improvement was also greater on two of the three remaining Accuracy scores, but the difference was not statistically significant. Contrary to what had been expected, the experimental group, which received training in interpersonal relations, did not improve more than the control group, which had not received such training. In fact, the trend of results suggests the opposite conclusion: that the training in interpersonal relations decreased accuracy.

Table 1 also presents the results for the Average-Variance scores. For each of the four measures the experimental-group change was significantly different from the control-group change. The experimental group in-

<sup>3</sup> For a discussion of importance of the manner of the selection of the objects in interpersonal perception studies see (3).

<sup>4</sup> For a detailed description of the measures and the scoring techniques see (2).



creased the average variance of their estimations more than did the control group from pre- to posttest. They exhibited more differentiation than did the control group; that is, they manifested less adherence to stereotype.

The results for the Average-Difference scores were similar. The experimental group changed more than did the control group in the amount of average difference that they saw between a patient's self-rating and where the patient "really was" on the personality scales. The results were statistically significant for both the first and second halves. Since the Ss in the experimental group increased their Average-Difference scores, they assumed less veridicality than did the control-group students.

These results indicate that the effect of the training program for the experimental group was to increase the amount of differentiation that the Ss made in their judgments about patients. There also appears to be a loss of accuracy, but the results were not conclusive.

#### *Relationship between Accuracy and Variability in Estimations*

The results obtained for the effect of training upon interpersonal perception were in accord with Cronbach's hypothesis that if an

TABLE 1

MEAN INCREASE, STANDARD DEVIATIONS AND *t* FOR DIFFERENCES BETWEEN GROUPS FOR INTERPERSONAL PERCEPTION SCORES

Measures	Experimental		Control		
	1st Half <i>N</i> = 20 2nd Half <i>N</i> = 17		1st Half <i>N</i> = 52 2nd Half <i>N</i> = 55		
	Mean Increase	<i>SD</i>	Mean Increase	<i>SD</i>	<i>t</i>
Accuracy scores					
Task A First half	1.70	5.95	3.88	5.92	1.36
Second half	-3.06	6.74	-3.89	6.17	0.46
Task B First half	-12.65	8.67	-6.81	2.78	2.46*
Second half	12.29	6.52	14.82	7.95	1.16
Average-Variance scores					
Task A First half	0.36	2.99	-1.51	3.43	2.08*
Second half	1.53	3.51	-0.29	2.50	2.30*
Task B First half	1.09	3.29	-1.57	3.47	2.85**
Second half	2.78	2.82	0.83	2.57	2.59*
Average-Difference scores					
First half	0.17	0.19	-0.06	0.26	3.48**
Second half	0.11	0.21	-0.06	0.21	2.78**

\*  $p < .05$ .

\*\*  $p < .01$ .

TABLE 2

CORRELATION COEFFICIENTS BETWEEN INCREASE IN ACCURACY, INCREASE IN AVERAGE-VARIANCE SCORES AND INCREASE IN AVERAGE-DIFFERENCE SCORES<sup>a</sup>

	Increase in Accuracy			
	Task A		Task B	
	1st Half	2nd Half	1st Half	2nd Half
Increase in Average-Variance scores				
Task A First half	-.19			
Second half		.13		
Task B First half			-.50*	
Second half				-.27*
Increase in Average-Difference scores				
First half	-.03		-.45*	
Second half		.07		-.26*

<sup>a</sup> *N* = 72.

\* Significant at .05 level.

*S* increases his differentiation among objects without a corresponding increase in estimating ability, he will make greater errors. But the findings reported apply to the groups of Ss only. A more direct way to study the relationship of variability and accuracy is to correlate the scores of the individual Ss on the two measures; therefore, change in Accuracy scores was correlated with the change in Average-Variance and Average-Difference scores for the first and second halves.

Table 2 presents the correlation coefficients obtained between change in variability and change in accuracy. Four of the eight correlations (all for Task B) were negative and significant beyond the .05 level of confidence. These results also support Cronbach's hypothesis.

TABLE 3

CORRELATION COEFFICIENTS BETWEEN ACCURACY AND AVERAGE-VARIANCE AND AVERAGE-DIFFERENCE SCORES FOR THREE ADMINISTRATIONS

	Accuracy	
	Task A	Task B
Average-Variance		
Pretest	.51	.57
Posttest I	.30	.76
Posttest II	.37	.55
Average-Difference		
Pretest	.34	.57
Posttest I	.35	.56
Posttest II	.33	.49

*N* = 72.

*r* = .28 at the .01 level.

Table 3 presents the correlations for each administration between Accuracy scores and both Average-Variance and Average-Difference scores. All of the twelve correlation coefficients were negative and significant beyond the .01 level of confidence. As the Ss made more variable estimations, their accuracy decreased. These results support quite conclusively the hypothesis that accuracy is negatively related to the amount of variability in Ss' estimations.

### DISCUSSION

Contrary to expectation, the group which received more training in interpersonal relations (the experimental group) did not improve in accuracy of interpersonal perception more than did the group without such special training (control group). In fact, the results support the opposite conclusion: that, as a result of their special training, the experimental group became less accurate than the control group.

The experimental group increased the variability in their estimations significantly more than did the control group. The experimental group also assumed significantly less agreement between the way Ss would rate themselves and the way they "really were." These results, in conjunction with the fact that the experimental group decreased in accuracy, support Cronbach's hypothesis that, if an S increases his differentiation without a corresponding increase in ability, he will make greater errors. This hypothesis was further supported by the finding that individual Accuracy scores were negatively correlated with both Average-Variance and Average-Difference scores. Additional (though less conclusive) support for the hypothesis stemmed from the finding that *change* in variability from pre- to posttest was negatively correlated with *change* in accuracy. The more that Ss increased the amount of variability they placed in their estimations, the less accurate they became. These results, taken together, constitute compelling evidence that the effect of the training program was to increase the variability that the trainees placed in their judgments and that their accuracy therefore decreased.

Investigation of the relationship between variability and accuracy in interpersonal perception has not progressed to the point

where Cronbach's hypothesis can be recommended as a basis for action. However, the implications of the hypothesis are so important that they deserve statement, particularly in regard to training programs such as the one reported in this study. These results indicate that training programs devoted to increasing accuracy of interpersonal perception (as in medical education and in the training of teachers, clinical psychologists, and similar specialists) run the risk of decreasing accuracy when they increase the trainees' responsiveness to individual differences. Since very little is known about how to train people to make more accurate judgments about others, training programs frequently utilize a procedure of "exposure" and little else. The belief that placing the trainee in a position to observe others and to make judgments will produce desirable results is challenged by these findings. Such experience may lead the trainee to differentiate among people far beyond his capacity to do so accurately. In the absence of dependable measures of his accuracy, the trainee lacks knowledge of his errors and may continue inappropriate overdifferentiation long after the training has ceased.

Cronbach has discussed this problem as follows: "Recognizing an optimum degree of differentiation makes it necessary to re-examine and qualify statements commonly made in training teachers, to the effect that every pupil has his own pattern and the teacher must fit methods to that pattern, not treat the pupil in terms of the statistical average" (1, p. 183). Whatever the ethical, moral, or ideological justification for emphasizing the unique characteristics of each person's individuality, the evidence so far indicates that "differentiation is harmful if the extent of adaptation or differentiation exceeds the amount justified by the accuracy of social perception" (1, p. 183).

### SUMMARY

The present study investigated the relationship between accuracy and variability in estimations and the effects of a training program in interpersonal relations upon interpersonal perception. Seventy-two Senior medical students were divided into an experimental group who received training in physician-patient relationships and a control



group who did not receive such training. At the beginning, during, and at the end of their Senior year, the students estimated the real status and self-ratings of patients presented in sound-film recorded interviews. Actual self-ratings and relevant MMPI scores were available for these patients as criteria.

Contrary to expectation, the experimental group tended to become less accurate than the control group. The experimental group increased the variability in their estimations significantly more than did the control group and also assumed significantly less agreement between the way subjects would rate themselves and the way they "really were." Accuracy scores were negatively correlated with variability scores and change in variability of estimation was negatively correlated with change in accuracy. These results, which agree with a hypothesis advanced by Cronbach, suggest that training programs devoted to increasing accuracy of interpersonal perception run the risk of decreasing accuracy when they

increase the trainee's responsiveness to individual differences.

#### REFERENCES

1. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177-193.
2. CROW, W. J. A methodological study of social perceptiveness. Unpublished doctoral dissertation, Univer. of Colorado, 1954.
3. CROW, W. J. The need for representative design in studies of interpersonal perception. *J. consult. Psychol.*, 1957, in press.
4. CROW, W. J., & HAMMOND, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, in press.
5. GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411-422.
6. HATHAWAY, S. R., & MEEHL, P. E. An atlas for the clinical use of the MMPI. Minneapolis: Univer. of Minnesota Press, 1951.
7. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951.

Received March 15, 1957.

# REFERENCE GROUPS, MEMBERSHIP GROUPS, AND ATTITUDE CHANGE<sup>1</sup>

ALBERTA ENGVALL SIEGEL<sup>2</sup>

*Stanford University*

AND SIDNEY SIEGEL<sup>2</sup>

*Center for Advanced Study in the Behavioral Sciences*

**I**N SOCIAL psychological theory, it has long been recognized that an individual's *membership groups* have an important influence on the values and attitudes he holds. More recently, attention has also been given to the influence of his *reference groups*: the groups in which he aspires to attain or maintain membership. In a given area, membership groups and reference groups may or may not be identical. They are identical when the person aspires to *maintain* membership in the group of which he is a part; they are disparate when the group in which the individual aspires to *attain* membership is one in which he is not a member. It has been widely asserted that both membership and reference groups affect the attitudes held by the individual (4).

The present study is an examination of the attitude changes which occur over time when reference groups and membership groups are identical and when they are disparate. The study takes advantage of a field experiment which occurred in the social context of the lives of the subjects, concerning events considered vital by them. The subjects were not aware that their membership and reference groups were of research interest; in fact, they did not know that the relevant information about these was available to the investigators.

The field experiment permitted a test of the general hypothesis that both the amount and the direction of a person's attitude change over time depends on the attitude norms of his membership group (whether or not that group is chosen by him) and on the attitude norms of his reference group.

<sup>1</sup> This study was supported by grants from the Committee for the Study of American Values at Stanford University and from the Stanford Value Theory Project. We wish to acknowledge with gratitude the assistance given by Davis W. Thompkins, Marilyn Sanchez-Corea, and Coleen Baker in the execution of this study, and the generous administrative cooperation of Elva Fay Brown, Dean of Women at Stanford University, and her staff.

<sup>2</sup> On leave from Pennsylvania State University.

This hypothesis is tested with subjects who shared a common reference group at the time of the initial assessment of attitudes. They were then randomly assigned to alternative membership groups, some being assigned to the chosen group and others to a nonchosen group. Attitudes were reassessed after a year of experience in these alternative membership groups with divergent attitude norms. During the course of the year, some subjects came to take the imposed (initially nonpreferred) membership group as their reference group. Attitude change after the year was examined in terms of the membership group and reference group identifications of the subjects at that time.

## THE FIELD EXPERIMENT

The Ss of this study were women students at a large private coeducational university. The study was initiated shortly before the end of their freshman year, when they all lived in the same large freshman dormitory to which they had been assigned upon entering the university. At this university, all women move to new housing for their sophomore year. Several types of housing are available to them: a large dormitory, a medium-sized dormitory, several very small houses which share common dining facilities, and a number of former sorority houses which have been operated by the university since sororities were banished from the campus. These latter are located among the fraternity houses on Fraternity Row, and are therefore known as "Row houses." Although the Row houses are lower in physical comfort than most of the other residences for women, students consider them higher in social status. This observation was confirmed by a poll of students (5, p. 205), in which over 90 per cent of the respondents stated that Row houses for women were higher in social status than non-Row houses, the remaining few disclaiming any information concerning status differences among women's residences.



In the Spring of each year, a "drawing" is held for housing for the subsequent year. All freshmen must participate in this drawing, and any other student who wishes to change her residence may participate. It is conducted by the office of the Dean of Women, in cooperation with woman student leaders. Any participant's ballot is understood to be secret. The woman uses the ballot to rank the houses in the order of her preference. After submitting this ballot, she draws a number from the hopper. The rank of that number determines the likelihood that her preference will be satisfied.

In research reported earlier (5), a random sample was drawn from the population of freshman women at this university, several tests were administered to the Ss in that sample, and (unknown to the Ss) their housing preferences for the forthcoming sophomore year were observed by the investigator. The Ss were characterized as "high status oriented" if they listed a Row house as their first choice, and were characterized as "low status oriented" if they listed a non-Row house as their first choice. The hypothesis under test, drawn from reference group theory and from theoretical formulations concerning authoritarianism, was that high status orientation is a correlate of authoritarianism. The hypothesis was confirmed: freshman women who listed a Row house as their first choice for residence scored significantly higher on the average in authoritarianism, as measured by the E-F scale (1, 2) than did women who listed a non-Row house as their first choice. The present study is a continuation of the one described, and uses as its Ss only those members of the original sample who were "high status oriented," i.e., preferred to live in a Row house for the sophomore year. In the initial study (5), of the 95 Ss whose housing choices were listed, 39 were "high status oriented," i.e., demonstrated that the Row was their reference group by giving a Row house as their first choice in the drawing. Of this group, 28 were available to serve as Ss for the follow-up or "change" study which is the topic of the present paper. These women form a homogeneous subsample in that at the conclusion of their freshman year they shared a common membership group (the freshman dormitory) and a common reference group (the Row). These Ss, however, had divergent

experiences during their sophomore year: nine were Row residents during that year (having drawn sufficiently small numbers in the housing drawing to enable them to be assigned to the group of their choice) and the other 19 lived in non-Row houses during that year (having drawn numbers too large to enable them to be assigned to the housing group of their choice).

E-F scores were obtained from each of the 28 Ss in the course of a large-scale testing program administered to most of the women students at the university. Anonymity was guaranteed to the Ss, but a coding procedure permitted the investigators to identify each respondent and thereby to isolate the Ss and compare each S's second E-F score with her first.

To prevent the Ss from knowing that they were participating in a follow-up study, several procedures were utilized: (a) many persons who had not served in the earlier study were included in the second sample, (b) the testing was introduced as being part of a nation-wide study to establish norms, (c) the test administrators were different persons from those who had administered the initial tests, (d) Ss who informed the test administrator that they had already taken the "Public Opinion Questionnaire" (E-F scale) were casually told that this did not disqualify them from participating in the current study.

The Ss had no hint that the research was in any way related to their housing arrangements. Testing was conducted in classrooms as well as in residences, and all procedures and instructions were specifically designed to avoid any arousal of the salience of the housing groups in the frame of reference of the research.

The annual housing drawing was conducted three weeks after the sophomore-year testing, and, as usual, each woman's housing ballot was understood to be secret. In this drawing, each S had the opportunity to change her membership group, although a residence move is not required at the end of the sophomore year as it is at the end of the freshman year. If an S participated in this drawing, the house which she listed as her first choice on the ballot was identified by the investigators as her reference group. If she did not, it was evident that the house in which she was currently a member was the one in which she chose to continue to live, i.e., was her reference group. With the information on each S's residence choice at the end of

her freshman year, her assigned residence for her sophomore year, and her residence choice at the end of her sophomore year, it was possible to classify the subjects in three categories:

- A. Women ( $n = 9$ ) who had gained assignment to live on the Row during their sophomore year and who did not attempt to draw out of the Row at the end of that year;
- B. Women ( $n = 11$ ) who had not gained assignment to a Row house for the sophomore year and who drew for a Row house again after living in a non-Row house during the sophomore year; and
- C. Women ( $n = 8$ ) who had not gained assignment to a Row house for the sophomore year, and who chose to remain in a non-Row house after living in one during the sophomore year.

For all three groups of Ss, as we have pointed out, membership group (freshman dormitory) and reference group (Row house) were common at the end of the freshman year. For Group A, membership and reference groups were identical throughout the sophomore year. For Group B, membership and reference groups were disparate throughout the sophomore year. For Group C, membership and reference groups were initially disparate during the sophomore year but became identical because of a change in reference groups.

As will be demonstrated, the Row and the non-Row social groups differ in attitude norms, with Row residents being generally more authoritarian than non-Row residents. From social psychological theory concerning the influence of group norms on individuals' attitudes, it would be predicted that the different group identifications during the sophomore year of the three groups of Ss would result in differential attitude change. Those who gained admittance to a Row house for the sophomore year (Group A) would be expected to show the least change in authoritarianism, for they spent that year in a social context which reinforced their initial attitudes. Group C Ss would be expected to show the greatest change in authoritarianism, a change associated not only with their membership in a group (the non-Row group) which is typically low in authoritarianism, but also with their shift in reference groups, from Row to non-Row, i.e., from a group normatively higher in authoritarianism to a group normatively lower. The extent of attitude change in the Ss in Group B would be expected to be intermediate, due to the conflicting influences of the imposed mem-

bership group (non-Row) and of the unchanged reference group (Row). The research hypothesis, then, is that between the time of the freshman-year testing and the sophomore-year testing, the extent of change in authoritarianism will be least in Group A, greater in Group B, and greatest in Group C. That is, in extent of attitude change,  $\text{Group A} < \text{Group B} < \text{Group C}$ .

RESULTS

*Group norms.* From the data collected in the large-scale testing program, it was possible to determine the group norms for authoritarian attitudes among the Row and the non-Row women at the university. The E-F scale was administered to all available Row residents ( $n = 303$ ) and to a random sample of residents of non-Row houses ( $n = 101$ ). These Ss were sophomores, juniors, and seniors. The mean E-F score of the Row women was 90, while the mean E-F score of the non-Row was 81. The E-F scores of the two groups were demonstrated to differ at the  $p < .001$  level ( $\chi^2 = 11.1$ ) by the median test (6, pp. 111-116), a non-parametric test, the data for which are shown in Table 1.

*Attitude change.* The central hypothesis of this study is that attitude change will occur differentially in Groups A, B, and C, and that it will occur in the direction which would be predicted from knowledge of the group norms among Row and non-Row residents in general. The 28 Ss of this study had a mean E-F score of 102 at the end of their freshman year. The data reported above concerning authoritarianism norms for all women residing on campus would lead to the prediction that in general the Ss would show a reduction in authoritarianism during the sophomore year but that this reduction would be differential in the three groups; from the knowledge that Row residents

TABLE 1  
FREQUENCIES OF E-F SCORES ABOVE AND BELOW COMMON MEDIAN FOR ROW AND NON-ROW RESIDENTS

	Residents of Non-Row Houses	Residents of Row Houses	Total
Above Median	36	166	202
Below Median	65	137	202
Total	101	303	404



TABLE 2  
FRESHMAN-YEAR AND SOPHOMORE-YEAR  
E-F SCORES OF SUBJECTS

Group	E-F Score		Difference
	End of Freshman Year	End of Sophomore Year	
A	108	125	-17
	70	78	-8
	106	107	-1
	92	92	0
	80	78	2
	104	102	2
	143	138	5
	110	92	18
B	114	80	34
	76	117	-41
	105	107	-2
	88	82	6
	109	97	12
	98	83	15
	112	94	18
	101	82	19
C	114	93	21
	104	81	23
	116	91	25
	101	74	27
	121	126	-5
	87	79	8
	105	95	10
	97	81	16
	96	78	18
	108	73	35
	114	77	37
	88	49	39

membership group identification and his reference group identification. The hypothesis that extent of attitude change would be different in the three subgroups of Ss, depending on their respective membership group and reference group identifications, is confirmed at the  $p < .025$  level; in extent of change in authoritarianism, Group A < Group B < Group C, as predicted.

Another way of looking at the data may serve to highlight the influence of membership groups and reference groups. At the end of the freshman year, the Ss in Groups A, B, and C shared the same membership group and the same reference group. During the sophomore year, the Ss in Group A shared one membership group while those in Groups B and C together shared another. From membership group theory, it would be predicted that the extent of attitude change would be greater among the latter Ss. This hypothesis is supported by the data (in Table 2): by the Mann-Whitney test (6, pp. 116-127), the change scores of these two sets of Ss (Group A versus Groups B and C together) differ in the predicted direction at the  $p < .025$  level. This finding illustrates the influence of *membership* groups on attitude change. On the other hand, at the conclusion of the sophomore year, the Ss in Groups A and B shared a common reference group while those in Group C had come to share another. From reference group theory, it would be predicted that attitude change would be more extensive among the subjects who had changed reference groups (Group C) than among those who had not. This hypothesis is also supported by the data (in Table 2): by the Mann-Whitney test, the change scores of these two sets of Ss (Groups A and B together versus Group C) differ in the predicted direction at the  $p < .05$  level. This finding illustrates the influence of *reference* groups on attitude change. Any inference from this mode of analysis (as contrasted with the main analysis of the data, by the Jonckheere test) must be qualified because of the nonindependence of the data on which the two Mann-Whitney tests are made, but it is mentioned here to clarify the role which membership and reference groups play in influencing attitude change.

The findings may also contribute to our understanding of processes affecting attitude change. The imposition of a membership

#### DISCUSSION

Substantively, the present study provides experimental verification of certain assertions in social group theory, demonstrating that attitude change over time is related to the group identification of the person—both his

generally are higher in authoritarianism than non-Row residents, the prediction based on social group theory would be that Group A would show the smallest reduction in authoritarianism scores, Group B would show a larger reduction, and Group C would show the largest reduction. The data which permit a test of this hypothesis are given in Table 2. The Jonckheere test (3), a nonparametric  $k$ -sample test which tests the null hypothesis that the three groups are from the same population against the alternative hypothesis that they are from different populations which are ordered in a specified way, was used with these data. By that test, the hypothesis is confirmed at the  $p < .025$  level.

group does have some effect on an individual's attitudes, even when the imposed group is not accepted by the individual as his reference group. This relationship is shown in the case of Group B. If the person comes to accept the imposed group as his reference group, as was the case with the Ss in Group C, then the change in his attitudes toward the level of the group norm is even more pronounced.

Methodologically, the study has certain features which may deserve brief mention. First, the study demonstrates that it is possible operationally to define the concept of reference group. The act of voting by secret ballot for the group in which one would like to live constitutes clear behavioral specification of one's reference group, and it is an act whose conceptual meaning can be so directly inferred that there is no problem of reliability of judgment in its categorization by the investigator. Second, the study demonstrates that a field study can be conducted which contains the critical feature of an experiment that is usually lacking in naturalistic situations: randomization. The determination of whether or not a woman student would be assigned to the living group of her choice was based on a random event: the size of the number she drew from the hopper. This fact satisfied the requirement that the treatment condition be randomized, and permitted sharper inferences than can usually be drawn from field studies. Third, the test behavior on which the conclusions of this study were based occurred in a context in which the salience of membership and reference groups was *not* aroused and in which no external sanctions from the relevant groups were operative. This feature of the design permitted the interpretation that the E-F scores represented the Ss' internalized attitudes (4, p. 218). Finally, the use of a paper-and-pencil measure of attitude and thus of attitude change, rather

than the use of some more behavioral measure, is a deficiency of the present study. Moreover, the measure which was used suffers from a well-known circularity, based on the occurrence of pseudo-low scores (1, p. 771; 5, pp. 221-222).

#### SUMMARY

In the social context of the lives of the subjects, and in a natural social experiment which provided randomization of the relevant condition effects, the influence of both membership and reference groups on attitude change was assessed. All subjects shared a common reference group at the start of the period of the study. When divergent membership groups with disparate attitude norms were socially imposed on the basis of a random event, attitude change in the subjects over time was a function of the normative attitudes of both imposed membership groups and the individuals' reference groups. The greatest attitude change occurred in subjects who came to take the imposed, initially nonpreferred, membership group as their reference group.

#### REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. GOUGH, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 237-246.
3. JONCKHEERE, A. R. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133-145.
4. SHERIF, M., & SHERIF, CAROLYN W. *Groups in harmony and tension*. New York: Harper, 1953.
5. SIEGEL, S. Certain determinants and correlates of authoritarianism. *Genet. Psychol. Monogr.*, 1954, **49**, 187-229.
6. SIEGEL, S. *Nonparametric statistics: For the behavioral sciences*. New York: McGraw-Hill, 1956.

Received February 27, 1957.



# SOCIOECONOMIC STATUS AND CONCEPTUAL THINKING<sup>1</sup>

JEROME SILLER

*Prosthetic Devices Study, New York University*

ALTHOUGH there is a considerable body of data indicating differences between socioeconomic groups on scores made on commonly used tests of intelligence, the meaning and significance of these differences is debatable. A better understanding of factors relating to socioeconomic group differences in mental abilities may be attained if we deal with separate phases of the intellectual process rather than with intelligence in its more global and ambiguous sense. The present concern is therefore with the ability to conceptualize and with its related process, concept formation. For this purpose, the following characterization of concept formation is adequate: "Behavior governed by concepts requires that the same response shall be made to objects of the same class, but a different response, or no response at all to objects not belonging to that class" (21, p. 609).

Davis and Havighurst (2, 3) have pointed out that status differentiations have the effect of varying degrees of defining and limiting the developmental environment of the child. They, among others, maintain that social class is an important factor in determining the development of meanings.

Eells (6), a member of this "culture fair" group, purportedly demonstrated the existence of a middle-class bias on intelligence tests, after an analysis of individual test items. Without necessarily accepting Eells' conclusion about test bias, there are two aspects of his data which are pertinent for the present study. Eells reported that mean status differences are largest for verbal and smallest for nonverbal material, and that the quality of attractiveness in the item appeared to produce large status differences. All differences were in the direction of high status superiority.

As in intelligence testing, there are virtually

no definitive experimental studies relating conceptual ability to socioeconomic status. Two investigations concern themselves with physical causality (4, 10). Both report little relationship between the scores on their concept tests and socioeconomic status. Each study, however, has methodological limitations, i.e., Huang's study (10) has many uninterpretable responses that may have influenced the analysis, while Deutsche's work has been severely criticized by Lacey (12). Finally, both are based upon a restricted aspect of conceptualization.

Dixon (5) reports significant differences on a test of contradictory relationships between children two to six years old when compared on the basis of parental occupation (children of students and professors vs. working mothers). Dixon explains these differences in terms of differential intelligence.

Findlay and McGuire (7), using a variation of the Vigotsky test, hypothesized that if cultural bias affects test performance, children from lower social status backgrounds might be expected to exhibit a higher degree of problem solving capacity (whatever that may be) than children from higher social status backgrounds in order to achieve the same range of scores in mental test performance as the children with obvious cultural advantages. Children from HS and LS backgrounds (grade levels 3, 6, and 9) were equated on IQ as measured by a group intelligence test and compared on the supposedly less culturally weighted conceptual test. The hypothesis was not validated, and differences were found in the opposite direction.

The above studies have demonstrated that the content of the test item is not the only factor in social status group differences. Another often mentioned factor is the form of symbolism (e.g., verbal or nonverbal) by which the item is presented. Eells' (6) findings in this respect have already been mentioned. Tyler (19) demonstrated that when Eells' data were converted into standard score form, the mean status differences between verbal and nonverbal items were no longer significant. Havighurst and Janke (8) and McGurk

<sup>1</sup>Based on a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in the Department of Psychology, Graduate School of Arts and Science, New York University. The author wishes to express his appreciation to Drs. J. V. J. Anna, Marie Jahoda, Elsa E. Robinson, and S. A. Liss. The valuable suggestions of Dr. Richard Christie throughout the study are deeply appreciated.

(16, 17) also report negative evidence in this respect. These studies indicate that it is by no means certain that class differences are more marked on verbal than nonverbal tests.

Conceptual ability in the abstract can only be expressed indirectly and through the media of certain forms of symbolism (verbal and nonverbal). The resultant group differences may not be differences in ability to conceptualize, as much as differences in proficiency (learned) with the type of symbolism used. The role of the supposed middle-class verbal superiority must thus be assessed. The initial purpose of this study was to determine whether status differences on conceptual tests are a function of the form of symbolism by means of which conceptual ability is being tested.

It has been noted that low status (LS) children tend to operate, as a group, on a more concrete level than do high status (HS) children (2, 14). Observation of the time perspective of LS and HS children indicates that the former are more restricted in this respect (13). Time perspective and sense of abstractness should be closely related, since a broad time perspective, by definition, corresponds with an abstract mode of response and a narrow time perspective with an immediate or concrete mode. This observation suggests that social status differences on tests of cognitive ability might also be a function of *actual* cognitive differences.

The present study therefore attempts an evaluation of the importance of the variables of form of symbolism and abstractness-concreteness for status differences on tests of conceptual ability.

The following hypotheses have been proposed as generalizations from findings in the area of intelligence testing to concept formation, and are tested in the present investigation.

Hypothesis 1. HS children score more highly than LS children on all tests of conceptual ability.

Hypothesis 2. The form of symbolism, verbal or nonverbal, of the conceptual test is a factor contributing to the status differences. Greater status differences are predicted on verbal material than on nonverbal material.

Hypothesis 3. When alternative answers are available which are alike in correctness but differ along the concrete-abstract dimension,

HS children more frequently select abstract answers than do LS children.

The foregoing hypotheses pertain to direct comparisons of the two groups wherein the subjects are *representative* of their particular status. The following hypotheses pertain to comparisons in which subjects from each group are matched on the basis of a specified variable.

Hypothesis 4. When subjects from the two groups are matched on the basis of nonverbal scores, HS children score more highly than LS children on tests of verbal concepts dealing with the same kind of conceptualization as the nonverbal tests.

Hypotheses 5. When subjects from the two groups are matched on the basis of scores on verbal concept tests, LS children score more highly than HS children on tests of nonverbal concepts dealing with the same kind of conceptualization as the verbal tests.

Hypotheses 4 and 5 are directed, in part, toward the problem of the hierarchical development of concepts as discussed by Heidebreder (9). It is assumed in Hypothesis 4 that LS and HS children equated on nonverbal material will draw apart, to the advantage of the latter, when verbal material is introduced. An explanation of the status differences in terms of verbal facility (Hypothesis 2) would not be contributive to the issue of innate group differences vs. differential and unequal group experiences. The latter position would argue that LS individuals simply have not had the opportunity to achieve verbal skills. Advocates of innate group differences might advance an alternate explanation in terms of hierarchical development. They might argue that verbal superiority is a natural outgrowth of greater intellectual ability in that verbal material is less "thing-like" than nonverbal material and therefore requires a higher order intellectual process.

Hypothesis 5 is a more direct check on the "hierarchy" principle. If one assumes, as did Findlay and McGuire (7), that greater overall ability is required by the LS child to do as well as the HS child on tests of verbal concepts, then, when nonverbal material is introduced, the LS child should be at an advantage.

Validation of the two hypotheses would serve to rule out hierarchical order of concept attainment as a factor in this study. Failure to



do so, however, would not serve as evidence for either position.

## METHOD

### Subjects

The Ss were 181 white sixth-grade New York City public school children whose average age was about eleven. They were divided into two groups, HS and LS. There were 99 in the HS group and 82 in the LS group. Only white middle-class children in a middle-class school and white lower-class children in a lower-class school were used. In all, ten classes from four schools were tested.

The average IQ of the two groups differed significantly at well beyond the .001 level. The intelligence test was the Otis QSMA, Form M. Summary data are presented in Table 1. Table 1 shows also that the HS group has a mean reading score of 8.03 on the Metropolitan Achievement Elementary Reading Test, a score two years greater than the mean score of the LS group, although the LS group is approximately at the reading level appropriate for its grade.

### Determination of Status

Status was primarily determined by free lunch data and a questionnaire filled out by the parents. The questionnaire, a modification of the Warner ISC (20), obtained information about education and occupation of both parents, number of people in the apartment and dwelling area. The 181 Ss finally used were from a larger pool of 309. The others were eliminated because of undetermined status, non-white, etc.

### Procedure

Each S was given the complete battery of tests, unless absent from one of the testing sessions. All of the tests were group pencil-and-paper examinations with generous time limits. None of the sessions exceeded sixty minutes. Each class was visited three times for testing.

### Description of Tests<sup>2</sup>

**Classification tests.** Two such tests were given: one verbal (VC) and the other non-verbal (NVC). The NVC test was part of the Lorge-Thorndike Intelligence Tests-Level 3, Form A (14). Three objects are shown to the left of the page. The S must pick one of the five objects to the right which best goes with

<sup>2</sup> Two additional tests were used in the original study (18): one was intended to be a nonverbal analogue of the Abstract-Concrete Similarities Test and the other a measure of status-oriented attitudes. The former proved to be technically inadequate and the latter is not relevant to the immediate problem. Discussion of these tests is therefore omitted here but may be found in the dissertation.

TABLE 1  
READING AND IQ CHARACTERISTICS OF SAMPLE

Variable	N	Mean	SD	Range	t
IQ					
High status	99	118.31	14.08	86-149	6.27**
Low status	78	102.65	17.82	69-142	
Reading					
High status	99	8.03	2.06	3.4-11.6	21.7**
Low status	76	5.97	1.97	3.0-11.6	

\*\* Significant beyond the .01 level.

the first three. The standardized time of nine minutes was allowed for the 25 items.

The VC test was specially constructed by the E from items reported in the literature and judged to be appropriate for the sixth grade. Attention was given to "culture fair" criticisms of item construction. The task of the S was to select the word that was most different from the remaining three. Ten minutes were allowed for the 35 items.

**Analogies tests.** Two analogies tests were given: one verbal (VA) and the other non-verbal (NVA). The NVA test was also part of the Lorge-Thorndike Intelligence Tests-Level 3, Form A (14). The ten minutes allowed for the 30 items was one minute more than the standard time allowance. The VA test was specially constructed by the E in a manner similar to the VC test. Testing time for the 35 items was 10'15".

The scoring for the classification and analogies tests was simply the number of correct answers.

**Abstract-Concrete Similarities (A-C Sim.)** A number of recent studies have successfully revealed qualitative differences on vocabulary tests. The procedure, in multiple-choice tests, has been to have all of the choices correct but ranging in quality from concrete to abstract. The task of the S is to pick what he believes to be the best answer of the ones offered. This procedure has served to demonstrate developmental differences (11) as well as normal-schizophrenic differences (1). The present administration is an attempt to extend its use to a single age group with the independent variable being social status.

The work of Chodorkoff and Mussen (1) served as a model for a specially constructed, objective vocabulary test suitable for sixth graders. The test, consisting of 35 words, was

to determine group preferences in type of definition selected. For each word, definitions in terms of class, description, example, function, and incorrect choice were constructed. The order of definitions was nonsystematic. A sample item follows: *Dog*: (a) it has four feet, a tail, and fur (description); (b) it means a pet dog (example); (c) it's a small animal (class); (d) something wooden (incorrect); (e) it barks and guards your house (function). No time limit was set. Each protocol was given five scores corresponding to the number of times that each type of definition was used.

RESULTS

Status Comparison: Unmatched Groups

*Hypothesis 1.* Testing of the first hypothesis involved the NVA, NVC, VA, and VC tests. Status comparisons were made by means of the Median test. Distribution free methods were used because most of the distributions were significantly skewed and remained so despite attempts at transformations. The skewness appears to be a result of excessive time limits.

The HS children did better than the LS children on all of the tests of conceptual ability, three of the four tests showing significant differences. Hypothesis 1 was thus substantiated. Summary data for the status comparisons are presented in Table 2.

*Hypothesis 2.* The hypothesis and accompanying prediction were borne out. As can be noted in Table 2, there were clearly significant differences between the status groups on the verbal tests. Only the NVC of the two nonverbal tests was found to have a median difference between groups great enough to be statistically significant at the .05 level, although the NVA was close to significance.

The absolute point differences between the

groups on the verbal tests were greater than the point differences on the nonverbal tests. Direct comparison, however, was not feasible as conversion to standard score units was not possible. However, findings discussed below in relation to the matched groups also support the hypothesis. To anticipate, no group differences were found when verbal scores were matched and comparisons made on nonverbal material, but, when the reverse was done, there were significant differences on the compared verbal material.

*Hypothesis 3.* Testing the third hypothesis involved the determination of which group selected more of each type of definition on the A-C Sim. test. Table 3 presents the group totals for each type of definition. Chi square was computed to test the hypothesis that choice of definition is independent of status. Since the chi square for 4 *df* was 102.05, significant at well beyond the .001 level, the hypothesis of independence could be rejected with confidence.

Examination of the contribution of each cell indicated that the class, example, and incorrect categories were solely responsible for the significant chi square. The actual difference, in terms of number of class responses, is small but statistically significant. There appears to be a greater proneness to use abstractions among the HS group than among the LS group.

A separate analysis of the importance of reading ability for choice of definition indicated that the two-year difference in group reading scores was not sufficient to explain the status differences. Proneness to respond in an abstract fashion was still more characteristic of the HS group than of the LS group even when the most inferior readers of the LS group were removed.

The significant difference should not obscure

TABLE 2  
NONPARAMETRIC MEDIAN TESTS OF DIFFERENCES  
BETWEEN THE HS AND LS GROUPS ON THE NVC,  
NVA, VA, AND VC TESTS

Test	Status Group with Larger Score	<i>p</i> *
NVC	High	.025-.05
NVA	High	.05-.10
VC	High	< .001
VA	High	.025-.05

\* Median test values were corrected for continuity.

TABLE 3  
COMPARISON OF THE LOW AND HIGH STATUS GROUPS  
IN PERCENTAGE SELECTING EACH TYPE OF  
DEFINITION  
(Total Number of Responses: HS = 3311, LS = 2642.  
Percentages Derived for Each Group Separately.)

Status	Descrip- tion	Class	Func- tion	Ex- ample	In- correct	Total
High	38.0	33.6	23.1	4.0	1.2	99.9
Low	36.4	27.9	23.6	9.2	3.0	100.1



the close similarity between the responses of the two groups. The type of choice is distributed in the same order with both groups having as an order of preference, description, class, function, example, and finally incorrect. It is questionable whether the group difference on abstraction is socially as well as statistically significant.

#### *Status Comparisons: Matched Groups*

Data bearing on Hypotheses 4 and 5 were obtained by matching members of each group on the basis of score on one form of symbolism and comparing them on the test of the same function using the other form of symbolism. Four such matchings were made using the classification and analogies tests. Analysis was by means of the Sign test, corrected for continuity.

*Hypothesis 4.* Two nonverbal matchings were made. NVC scores were equated and VC scores compared, and NVA scores equated and VA scores compared. Both matchings showed a significant difference in favor of the HS group. Hypothesis 4 was confirmed. The data are summarized in Table 4.

*Hypothesis 5.* Two verbal matchings were made. VC scores were equated and NVC scores compared, and VA scores were equated and NVA scores compared. The data, as shown in Table 4, indicate that the hypothesis was not confirmed by either of the matchings. It was not possible to assess the importance of order of attainment.

*Excluded cases.* The number of matchings for each test was always less than the total number of cases involved for each of the tests. Analysis of the equated groups indicated that they were not representative of their status levels. They were composed primarily of the lower and middle range of scores of the HS population and the middle and higher range of scores of the LS population. It was possible to equate many Ss in the upper range of scores; however, the HS Ss in this range could not be completely matched by the LS Ss since there were more of the former in the sample. LS and HS Ss could not be matched on the lower range of scores to any appreciable degree because there were very few HS Ss who scored very low on the various tests while there were many LS Ss who did so. The excluded Ss represent a one-tailed extreme for each of the

TABLE 4  
NONPARAMETRIC SIGN TESTS OF DIFFERENCES  
BETWEEN MATCHED GROUPS ON THE NVC, NVA,  
VA, AND VC TESTS

Test Matched	Test Compared	N	HS <sup>a</sup>	LS <sup>a</sup>	Ties	p <sup>*</sup>
NVC	VC	68	45	22	1	.007
NVA	VA	60	39	21	0	.028
VC	NVC	56	22	32	2	—
VA	NVA	55	28	27	0	—

\* Sign test values were corrected for continuity.

<sup>a</sup> Number of times member of that status exceeds matched mate.

status groups, the implications of which fact will be considered below.

#### *Status Comparisons Matched for Intelligence*

Fifty Ss from each group were closely matched for intelligence. The mean IQ for the LS group was 111.96 and 112.00 for the HS group. All 28 of the excluded LS Ss had IQs below 100. All but one of the excluded HS Ss had IQs over 100. Again, we note that the equated groups were not unbiased samples of their parent population. There were no significant differences between the equated groups when comparisons by means of the Sign test were made on the NVC, VC, NVA, and VA tests. Unlike the findings of Findlay and McGuire (10) who used IQs ranging only from 90-110, the direction of differences did not favor either status.

#### DISCUSSION

So far, emphasis has been on defining the functional importance of the variables, form of symbolism and proneness to use abstractions, for a generalized consideration of conceptualization. This approach is in contrast to the few previous studies in the area which have generally considered restricted aspects of conceptualization.

Examination of the "one-tailed status extremes" excluded in matching facilitated insight into the meaning and possible basis of the status differences. The excluded LS Ss, it will be remembered, were from the low range of scores in the LS population and the excluded HS Ss from the high range of scores in the HS population. Only because of the apparently strong relationship between verbal performance and status were significant differences found for some of the equated

groups. On the other hand, when reading score was equated, as was done on the A-C Sim. test, and, in effect, only the middle range of Ss compared, no over-all status differences were found in type of definition. Again, there were no significant differences when the groups were based on IQ. It was obvious from an inspection of the frequency distributions in all of these instances that important segments of each of the status groups (the one-tailed status extremes) and most particularly the LS, were not entering into the group comparisons.

A fraction of the LS group consistently scored low on most or all of the tests. Placement in the bottom quartile on three or more tests was arbitrarily fixed as the criterion for terming an S a "low scorer." There were 19 such individuals of the 82 in the LS group. There was no subgroup corresponding to the "low scorers" in the HS population although about a half-dozen individuals did score consistently low on more than two of the tests. However, even these individuals had scores which generally were higher than that of many of the LS "low scorers."

Having isolated this "low scorer" group it was possible to evaluate their role in the production of the differences between the HS and LS groups. Statistical comparison of the groups, with only the "low scorers" omitted, yield similar medians, distributions, etc. Comparison of the frequency distributions clearly indicates that the statistically significant group differences arose much more out of the infrequency of low scoring HS Ss than from the lack of high scoring LS Ss. For the most part, the two status samples are very similar in test performance, and the status differences are a function of a relatively small number of Ss in the LS group.

On the basis of the above analyses, it appears to be misleading to refer to general status differences in conceptual ability, even though such differences can be statistically derived. A more accurate statement of the situation as it appears from the present study would be that when one considers status differences in conceptual ability there is a greater possibility of finding Ss with low ability in an LS group than in an HS group but that, for the most part, performance of LS and HS Ss is similar.

Why should there be a prevalence of "low

scorers" in the LS sample and inferentially in the LS population? It is possible that an answer to this question may shed new light in the recurrent controversy over the source of status differences in cognitive ability. Although the study was not designed to provide such an answer, it does suggest a hypothesis which might contribute to an answer: that the "low scorers" might be distinguished from the other LS Ss in that the former come from families that are not oriented in terms of upward social mobility while the latter come from upwardly mobile homes. As a concomitant of this attitude toward social mobility we might expect greater psychological impoverishment in the daily existence of "low scorers" than in the other LS Ss. Status in the present study was defined in objective sociological terms and not as the families considered themselves. It is likely that Ss from upwardly mobile LS families have psychological environments more similar to HS families than to sociologically determined LS families who are not so oriented.

Social immobility, of course, is not the exclusive or even necessarily the most important factor to consider in reference to the "low scorer" group. There unquestionably are some children in the group with inherently low intellectual endowment, as there are others who have been retarded by experiential impoverishment. Other factors to be evaluated are prenatal environment, nutritive deficiencies, and, in this particular sample, a greater proportion of first generation Americans in the LS than in the HS group.

It is obvious that the necessary data for explaining the basis for status differences will not be derived from the psychometric type of study (such as this one) which ends by showing differences, when an explanatory study should begin from this point. However, if other studies show effects similar to that of the "low scorer" group in influencing the results and if provision is made to analyze the characteristics of these Ss and their families, the meaning of average status differences may be clarified.

#### SUMMARY

This study was designed to relate socioeconomic status to conceptual ability. The basic problem was to evaluate the importance of (a) the form of symbolism by means of



which conceptual ability was being tested and (b) the abstractness of the material for status differences in performance.

A measure of status attitudes and four conceptual tests were administered to groups of 99 high status and 82 low status white sixth-grade public school children.

The major results were as follows:

1. As groups, HS children did better than LS children on all tests of conceptual ability, particularly those involving verbal material.
2. The HS group selected more definitions of an abstract type than did the LS group.
3. The status groups were matched for a number of variables such as verbal score, nonverbal score, and intelligence, and then compared on a number of other variables. Generally there were no significant status differences. This procedure was used to determine whether the hierarchical order of concept attainment principle could be ruled out as a factor in HS verbal superiority. The results were inconclusive.
4. The matchings usually excluded low scoring Ss in the LS group and high scoring Ss from the HS group. When the relative small number of lowest scorers in the LS group was removed from the group comparisons there were no status differences. The meaning of general group status differences was questioned and further study of the "low scorer" group strongly suggested.

#### REFERENCES

1. CHODORKOFF, B., & MUSSEN, P. Qualitative aspects of the vocabulary responses of normals and schizophrenics. *J. consult. Psychol.*, 1952, **16**, 43-48.
2. DAVIS, A. *Social-class influences upon learning*. Cambridge, Mass.: Harvard Univer. Press, 1948.
3. DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child-rearing. *Amer. sociol. Rev.*, 1946, **11**, 698-710.
4. DEUTSCHE, JEAN M. The development of the children's concepts of causal relations. In Barker-Kounin-Wright (Eds.), *Child behavior and devel-*

*opment*. New York: McGraw-Hill, 1943, Ch. VIII.

5. DIXON, J. C. Concept formation and emergence of contradictory relations. *J. exp. Psychol.*, 1949, **39**, 144-149.
6. EELLS, K., DAVIS, A., HAVIGHURST, R. J., HERICK, V. E., & TYLER, R. *Intelligence and cultural differences*. Chicago: Univer. of Chicago Press, 1951.
7. FINDLAY, D. C., & MCGUIRE, C. Social status and abstract behavior. *J. abnorm. soc. Psychol.*, 1957, **54**, 135-137.
8. HAVIGHURST, R. J., & JANKE, L. L. Relations between ability and social status in a midwestern community: I. Ten-year-old children. *J. educ. Psychol.*, 1944, **35**, 357-368.
9. HEIDBREDER, EDNA. The attainment of concepts: I. Terminology and methodology. *J. gen. Psychol.*, 1946, **36**, 173-189.
10. HUANG, I. Children's conception of physical causality; a critical summary. *J. genet. Psychol.*, 1943, **63**, 71-121.
11. KRUGLOV, LORRAINE P. Qualitative differences in the vocabulary choices of children as revealed in a multiple choice test. *J. educ. Psychol.*, **44**, 1953, 229-243.
12. LACEY, J. I. Review of Deutsche (see 7). *Amer. J. Psychol.*, 1939, **52**, 318-320.
13. LESHAN, L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, **47**, 589-592.
14. LORGE, I., & THORNDIKE, R. L. *The Lorge-Thorndike Intelligence Tests*. Boston: Houghton Mifflin, 1954.
15. MCCANDLESS, B. R. Environment and intelligence. *Amer. J. ment. Def.*, 1952, **56**, 674-691.
16. MCGURK, F. C. J. Socioeconomic status and culturally weighted test scores of Negro subjects. *J. appl. Psychol.*, 1953, **37**, 276-277.
17. MCGURK, F. C. J. On White and Negro test performance and socioeconomic factors. *J. abnorm. soc. Psychol.*, 1953, **48**, 448-50.
18. SILLER, J. The effect of differential socioeconomic status upon the development of conceptual ability. Unpublished doctoral dissertation, New York Univer., 1956.
19. TYLER, F. T. Comments on the correlational analysis reported in intelligence and cultural differences. *J. educ. Psychol.*, 1953, **44**, 228-295.
20. WARNER, W. L., MEEKER, M., & EELLS, K. *Social class in America*. Chicago: Science Research, 1949.
21. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

Received March 4, 1957.

# RECALL OF INTERRUPTED TASKS UNDER STRESS: A PHENOMENON OF MEMORY OR OF LEARNING?<sup>1</sup>

ALBERT J. CARON AND MICHAEL A. WALLACH

*Harvard University*

**P**ERHAPS no experimental approach to repression has received more attention than the attempt to administer Zeigarnik's interrupted-task procedure (22) under ego-involving conditions. The rationale underlying this use of the Zeigarnik technique held that ego involvement would cause incompleted and completed tasks to be viewed as failures and successes respectively, and hence, in accordance with repression theory, would lead to the expulsion of incompleted items from consciousness. The prediction followed, therefore, that the typical "Zeigarnik" recall pattern (superior recall of incompleted tasks) would be replaced by a "repressive" recall pattern (superior recall of completed tasks).

Since this view was first advanced, much evidence has accumulated (1, 2, 3, 7, 8, 18, 19) which indicates, not surprisingly, that such a reversal of the Zeigarnik ratio is not a function of stress alone, but depends, in addition, on personality factors—some individuals recalling a preponderance of successes and others yielding extreme Zeigarnik ratios under ego-involving conditions. While the isolation of the particular personality dimensions associated with each of these stress recall patterns offers a fertile field for personality research, the mechanisms assumed to underlie each pattern—namely, repression of failures in the case of selective recall of successes (S-recall) and superior trace stabilization of failures in the case of selective recall of failures (F-recall)—are still open to question. As others (15, 20) have observed, finding a selective recall preference for success items under stress is not definitive proof that failure items were repressed. An alternative explanation, selective *learning* in favor of success items, has not been ruled out. Likewise, a selective recall preference for failure items under stress may arise not from enhanced stabilization of their traces but rather from their being better learned in the

first place. The obvious experimental solution here—to equate successes and failures for original learning—is hard to achieve, for if failure items are really threatening, then the same factors which theoretically lead to their repression (or trace stabilization) may also be expected to impair (or improve) their learning.

The aim in the present study was to provide a more feasible test of the memory versus learning explanations of both these recall patterns. The approach employed follows from an analysis of the assumptions underlying interpretation of each pattern as a memory or learning effect. Consider first the alternative assumptions with respect to S-recall under stress.

As has been noted (4, 23), the essential postulate of repression theory is that threatening events are not actually forgotten but persist in an unconscious state, continuously striving to regain consciousness. Moreover, should the threatening character of repressed events be allayed (via psychotherapy or other means), they should *re-emerge* into consciousness. The selective learning position, on the other hand, implies no such restoration for forgotten items: it assumes that decreased recall results entirely from a deficiency in original registration. The two interpretations thus offer different predictions for the recall of events whose threatening nature is allayed after learning—the repression view implying enhanced recall, the selective learning view implying no change in recall, for such items. In the present study, two groups (both known to exhibit S-recall under stress) were subjected to a stressful interrupted-task procedure. For one group, however, recall was preceded by a quasi-therapeutic session in which the test was exposed as an experimental hoax. If the S-recall pattern is indeed due to repression, such exposure, by reducing associated anxiety, should restore the recall of incompleted items, and hence cause this group to shift toward a Zeigarnik pattern. If, on the other hand, selective learning underlies S-recall, then no such shift should occur.

With respect to the F-recall pattern, ex-

<sup>1</sup> This research was conducted while the senior author held a U.S.P.H.S. postdoctoral research fellowship at Harvard University. Grateful acknowledgment is made to Dr. Jerome S. Bruner for making available the facilities of his laboratory.



planation in terms of selective learning is much like that offered above for S-recall. Recall is said to be totally dependent on learning and, whatever the reason may be, failure items have registered more strongly than success items. The interpretation in terms of selective remembering, on the other hand, is based, according to Rosenzweig's recent formulation (17), on the task-tension theory of Lewin (12) and Zeigarnik (22). According to this view, the traces of failure items are better stabilized than those of success items, because the former are laid down in a region of high psychic tension—tension aroused by instructions to perform a task and dischargeable only by its completion. Rosenzweig has called this a "need-persistent" reaction to failure as contrasted with the "ego-defensive" reaction of repression. This position predicts that anything done after learning to reduce task tension should bring about a reduction in the recall of failure items. The selective learning view, on the other hand, implies no recall change as a consequence of such post-learning manipulation. To test these predictions, two other groups—both known to exhibit F-recall under stress—were subjected to the same stressful situation described above, and recall for one group again followed exposure of the experimental hoax. If the F-recall pattern is indeed due to superior trace stabilization of failures, the revelation that incompletable tasks were actually impossible to solve should serve to discharge any residual tension associated with incompletable and thereby tend to reduce the recall of failures. No such reduction in recall of failures should occur, on the other hand, if this effect is due to differential learning.

A final question examined in the present study may now be mentioned. If our results should indicate that both the S- and F-recall patterns are a function of learning rather than retention, does this mean that there is no evidence in the present experiment for the operation of mnemonic processes? This question is prompted by consideration of an aspect of the present kind of procedure which distinguishes it from the original Zeigarnik technique, a difference heretofore neglected by most writers. To construct a situation sufficiently threatening for college students, the experimental tasks have usually been presented in the guise of an intelligence test and indeed are so presented here. In so doing, however,

the procedure is given a unity and integration which is not present in the typical Zeigarnik situation, where *S* is merely asked to perform a series of disparate and ostensibly unrelated tasks. Under intelligence-test conditions, then, it is entirely possible that *S*'s defensive reaction is oriented more toward the test as a whole, i.e., toward the single experience of examination failure, rather than toward particular items within the test, and hence selective recall of individual items may not be a measure optimally sensitive to *S*'s invoked defense. Memory for the entire test situation would seem more pertinent, but for obvious reasons such a measure is unfeasible. To determine whether there was indeed a mnemonic reaction to the total test situation in this study, an alternative indicator in the area of perception was employed. The rationale for its use is as follows: If *S* tends to repress an experience, aspects of that experience should be relatively inaccessible as hypotheses for perceptual recognition. Likewise, if the trace of an experience is well stabilized for *S*, aspects of it should be hyperaccessible as recognition hypotheses. Thus, following the stress situation, words related to it in a general way were presented tachistoscopically (along with matched neutral words). A repressive reaction to the entire situation should result in heightened recognition thresholds for such words; enhanced trace stabilization, on the other hand, in lowered thresholds. Moreover, if these effects are truly memorial, and not due to registration processes, they should cease to occur when exposure of the experimental hoax precedes the perceptual test.

## METHOD

### Subjects

Since none of the personality or performance measures previously shown to relate to selective recall under stress have been conclusively validated, it was not considered appropriate to use any one or combination of them as a basis for preselecting the S-recallers and F-recallers in our various treatment groups. As an alternative to preselection, it was decided to administer a large battery of promising measures<sup>2</sup> to randomly selected *S*s already tested in our experimental conditions, to factor analyze these measures, and, from the correlations between the resulting factors and the

<sup>2</sup> The measures used included tests related rationally as well as experimentally to each of three personality areas: achievement motivation, hysteria, and ego strength.

selective recall scores of our stress group, to isolate a stable predictor that could be used *post hoc* to subdivide all treatment groups. This analysis yielded four orthogonal factors, only one of which was at all related to recall direction in the stress condition. The measure with the highest loading in this factor (and, incidentally, with negligible loading in the other factors) was the dichotomous variable of private prep school versus public high school attendance, and indeed this single variable was more strongly associated with our recall measure<sup>3</sup> than any other variable or combination of variables defining this factor (see Footnote 4 for the major variables). Thus, the public-private dichotomy involved fewer dimensions than one might at first have supposed, and also accounted for most of the factorial relationship with recall direction. These considerations, coupled with considerable evidence suggesting that the reactions of Harvard public and private school Ss should be quite different in an intelligence-test situation,<sup>4</sup> led us to use this variable to subdivide our groups into F-recaller and E-recaller components.

The Ss were chosen in the following way: Invitations to participate in the experiment were mailed out to every tenth name in the list of Harvard freshmen. The list was gone through twice in this manner, starting at a different name each time. The letter attractively described the experiment as a study of "symbolic processes," and also emphasized the remuneration for participation. About half of the men to whom letters were mailed volunteered—125 in all, of whom 117 completed the experiment—and they were randomly assigned to the several experimental conditions. The number of public and private school Ss in each condition was roughly equivalent.

### Sequence of Procedures

The order of procedures for all Ss was as follows: (a) initial digit recall test (4 minutes);

<sup>3</sup> Chi square was significant at less than the .025 level with a one-tail test—use of a one-tail test being justified since recall directions were predicted from the findings discussed in Footnote 4.

<sup>4</sup> McArthur's studies (13, 14) of the personalities of Harvard public and private school boys, for instance, indicate that the former tend to have a stronger drive for achievement and to be more sure of their personal and intellectual worth than the latter. Research by Atkinson (3) provides further support for this view. Ss with high need for achievement recalled more failures under stress and more successes under neutral conditions, while Ss with low need for achievement did the opposite—results which directly parallel our recall findings for public and private school Ss respectively. Finally, we may consider the performance of public and private school boys on some of our independent measures. The public school students made significantly higher grades at Harvard than those from private schools ( $p < .02$  with a two-tail test), although the two groups did not differ in intelligence. Moreover, on the factor mentioned above, public school attendance was associated with such achievement variables as level of aspiration, while private school attendance was related to such indications of personal insecurity as the Knutson Personal Security Inventory (11) and the Social Introversion scale of the Guilford STDCR battery (10).

(b) scrambled sentence tasks (50 minutes); (c) second digit recall test (4 minutes); (d) interim discussion with E (8–10 minutes); (e) initial recall of scrambled sentences (5 minutes); (f) determination of perceptual recognition thresholds (60 minutes); (g) second recall of scrambled sentences (5 minutes); (h) third recall of scrambled sentences, two days later (5 minutes).

### Tests

*Digit recall.* Half of the Wechsler-Bellevue Digit Recall Test (21) was administered before, and half after, the scrambled sentence tasks, one instance of each forward and backward couplet being presented in each administration. All responses were written from left to right, and writing never began until E finished reading the digits. The number of digit sequences completely correct provided the metric. Scores for the first half alone constituted a check on the equality of memory ability for the several experimental groups; the difference between scores for the first and second halves provided a measure of the anxiety induced by the intervening work on scrambled sentences.<sup>5</sup>

*Scrambled sentences.* The experimental tasks were sixteen 20-word sentences similar to ones used by Alper (1), Eriksen (6), and Caron (5). Each sentence was divided into 8 to 10 two- and three-word phrases and was presented in scrambled form, S's job being to rearrange them into a meaningful sentence. Half of the sentences, randomly distributed in the series of sixteen, were unsolvable. Contents of the sentences were diversified so that recall of one would be no aid to recall of another, and the completed and incomplete sentences were equated for recall difficulty in a previous study [see Caron (5)]. To facilitate group administration, the sentences were presented in booklet form, one sentence to a page. A  $2\frac{1}{4}$ -min. time limit was set for each solution and all Ss were required to start a new sentence at the same time. As an aid to recall, each sentence was named by selecting one phrase that best represented its meaning and placing this phrase above the scrambled sentence in the test booklet. E called out this "name" phrase at the start of each problem. For both initial and later recall tests, S had no prior knowledge that recall would be requested. In recalling sentences, S was required to write the name of a sentence or any other key phrase or word sufficient to identify it without question, no more than five minutes being allowed for each recall session. The measure of recall direction was the number of completed minus the number of incomplete tasks recalled (CR-IR).

*Perceptual materials.* Immediately after initial recall, perceptual recognition thresholds were determined for words pertaining in a general way to the scrambled sentence test and for neutral words of matched structure and of matched frequency (according to the Thorndike-Lorge L count). The former group included the following words: *sentence, booklet, phrases, examination, reasoning, scrambled, and tested*; the corresponding neutral words were *sunshine, cookies, sketches, agriculture,*

<sup>5</sup> For the use of digit recall as a measure of anxiety, see Rapaport *et al.* (16).



*leisurely, channeled, and regard.* Ss, tested in groups of four, were seated five feet from a ground glass screen in a room illuminated solely by a 40-watt fluorescent ceiling bulb one foot behind S. Each word, no more than 14 inches in length, was flashed at  $\frac{1}{100}$ -second, beginning at subthreshold brightness levels and rising in two-volt steps until all Ss recognized the word. Threshold was taken as the voltage level at which correct report first occurred. The order of threatening and neutral words was random, with matched words always being seven positions apart. To prevent communication between Ss, responses were written, E observing these and telling each S to stop and turn over his paper when he had written the correct word.

### Conditions

*Stress.* The stress condition was structured with two objectives in mind: (a) to convince S that he was taking a highly valid and discriminating intelligence test, and (b) to make him feel that he had done very poorly on it. With respect to the former objective, a number of techniques were used: In the first place, the scrambled sentence task was introduced as a test recently developed by the "Stanford people" in connection with a top level government project, a brief comment being added on the relationship of the Stanford people to the Stanford-Binet test. Second, E implied that the government project involved careful selection of personnel for critical intelligence work, and hence that the present test had been designed to provide precise discrimination in the average to superior range of adult intelligence. Third, bogus validity data were presented specifying academic and professional success as criteria. Finally, the title "Stanford Reasoning Test" was printed on the cover of the test booklet, and beneath that were printed "name," "age," and "IQ." (IQ was to be filled in, if known. Accomplices, discussed below, always filled in high IQs.) That an intelligence test would be administered in a study ostensibly concerned with "symbolic processes" was rationalized on the ground that E wished to compare the symbolic performances of average and highly intelligent students.

The following means were used to make S feel that he had done badly on the test: First, E stated that most Harvard students should perform very well on this test, since they were undoubtedly at the high end of the distribution of adult intelligence. Second, Ss were tested in groups of six, and each such group included two accomplices who pretended to solve all the tasks rapidly and exuded an air of confidence throughout the session. Finally, the four naive Ss received alternate forms of the test such that when any two of them failed to complete a task the other two Ss completed that same task, the result being that for any single task, four of the six testees achieved solution while two did not.

After completion of the exam, the second digit recall test was administered, and was immediately followed by a period of 8-10 minutes during the first part of which E made remarks intended to reactivate feelings of stress (referring to the exam's highly discriminative nature, the fact that scores would be made available, etc.), and during the remainder of which he scheduled Ss for subsequent sessions. Such scheduling was done at this point in order to make the interim period for the stress condition as long as that for the relief condition, and yet prevent Ss in the stress condition from discuss-

ing the exam among themselves. Half of the stress Ss were kept under stress for all three recall sessions, while the other half were relieved just prior to the second recall test.

*Relief.* This condition was exactly the same as the stress condition, except for the contents of the 8-10-minute interim period mentioned above. Rather than reactivate feelings of stress during this period, E attempted to relieve such feelings. He said the following: "Everybody sit back and relax. From the looks on your faces I gather that this was a pretty rough test—but perhaps you didn't do as badly as you think. You'll hate me when I tell you this, but believe me, it was all in the interest of science..." E then explained that the entire situation had been a hoax—that half the tasks had been unsolvable, that the two apparent geniuses in the group had really been E's accomplices, and so on. This revelation released a flood of affect—best described as a mixture of surprise and tremendous relief—which was inevitably followed by a round-robin of descriptions of one's feelings and reactions during the session. The subsequent airing of these feelings gave the session a definite therapeutic tone.

*Neutral.* In contrast to the stress condition, the aim of the neutral condition was to create as informal and relaxed an atmosphere as possible, and to focus attention on the tasks rather than on S's performance. To this end, the scrambled sentence tasks were introduced as a pretest of materials for a future experiment. The fact that the tasks were timed was rationalized by telling S that the purpose of the pretest was to eliminate tasks requiring too much time, E supposedly being interested in tasks that took no longer than about two minutes to solve. Moreover, whereas the timing for the stress condition was done in a test-like manner with E holding his stop watch conspicuously, the timing in the neutral condition was done unobtrusively and casually. The booklet of scrambled sentences, in addition, had no title on it, and S was told that he need not supply his name. Finally, no accomplices were present and all Ss had the same form of the test, so that everyone completed, or failed to complete, any given task. The 8-10-minute interim period between the second digit recall test and initial recall of the scrambled sentences was spent scheduling Ss for further sessions and having them fill out parts of some questionnaires.

### RESULTS

Before considering the results which bear directly on our major questions, let us examine the data pertaining to a basic assumption of the study; namely, that the cognitive effects obtained in our stress condition were due to the ego-involving operations themselves, rather than to something inherent in the test materials or to subject differences alone. Comparison of the stress and neutral conditions for each of the subject groups indicates that this assumption is entirely tenable: Private and public school groups yielded significant differences ( $p < .01$  in each case) between stress and neutral conditions on both the recall difference (Table 1, Column 2) and the per-

TABLE 1  
MEAN INITIAL DIGIT RECALL (D), SCRAMBLED SENTENCE RECALL SCORES, AND PERCEPTUAL RECOGNITION SCORES FOR S-RECALLERS (PRIVATE SCHOOL Ss) AND F-RECALLERS (PUBLIC SCHOOL Ss) UNDER THREE EXPERIMENTAL CONDITIONS<sup>a</sup>

Condition	D	CR-IR	TR	CR	IR	P <sup>b</sup>
S-recallers (Private School Ss)						
Stress	9.5 (N = 19)	+1.1	5.5	3.3	2.2	+6.0
Relief	9.3 (N = 19)	+0.6	4.4	2.5	1.9	+0.9
Neutral	8.6 (N = 15)	-0.2	6.4	3.1	3.3	-1.7
F-recallers (Public School Ss)						
Stress	9.0 (N = 23)	+0.1	5.3	2.7	2.6	-2.6
Relief	10.2 (N = 23)	+0.1	5.7	2.9	2.8	-0.7
Neutral	9.4 (N = 18)	+1.2	6.8	4.0	2.8	+3.8

<sup>a</sup> The scrambled sentence recall and perceptual recognition measures include mean difference between number of completed and incompleting tasks recalled (CR-IR), mean total number of tasks recalled (TR), mean number of completed (CR) and incompleting (IR) tasks recalled, and mean perceptual threshold difference for threat-related and matched neutral words (P).

<sup>b</sup> A positive difference means higher thresholds for threat-related than for neutral words. The brightness difference in volts is twice the size of each entry.

ceptual threshold scores (Table 1, Column 6). Moreover, the directions across recall and perception were completely consistent for each group, private school Ss exhibiting S-recall as well as "repressive" perceptual patterns under stress but just the opposite tendencies under neutral conditions, public school Ss doing exactly the reverse. (The interaction *F*s are significant beyond the .01 level for each variable. See Table 2.)<sup>6,7</sup>

<sup>6</sup> Lack of a strong recall tendency in favor of incompleting tasks in any of our subgroups indicates the presence of a general recall bias in favor of completed tasks. Thus, when the term "F-recall pattern" is employed in the present study, it is being used in a relative, not an absolute, sense. That this term is nevertheless being used appropriately, i.e., to specify a type of cognitive reaction, is indicated by the finding, noted earlier, that our group differences directly parallel those of Atkinson (3) for high and low need achievers tested under similar conditions.

<sup>7</sup> In all our analyses of variance, the form of the test,

For both private and public school Ss, total recall drops significantly from neutral to stress conditions (Table 1, Column 3). However, while for private school Ss this is due entirely to a decline in recall of failures (Column 5)<sup>8</sup> for public school Ss it is completely determined by a decline in recall of successes (Column 4). These losses assume added import when one compares the groups on our measure of recall ability—digit-retention. Column 1 of Table 1 shows the mean pretest digit-retention scores for the subgroups. The analysis of variance of these scores yielded no significant or near significant *F*s (Table 2). Consider now the findings that indicate whether these losses are due to mnemonic or learning processes.

1. *Is the S-recall pattern a function of repression of failures or selective learning in favor of successes?* The critical datum here is the degree of shift in the recall difference score from stress to relief conditions for private school boys (Table 1, Column 2). Although a slight shift occurred, it was far from significant, *t* being less than 1. Recall of incompleting tasks (Table 1, Column 5) shows a nonsignificant decrease from stress to relief conditions, whereas, if a repression effect were operating and given the mnemonic comparability of our groups, one would expect, if anything, an increase (memory for failures being restored by the relief treatment). Recall of completed tasks (Column 4) and total recall (Column 3) also exhibit nonsignificant declines. In sum, then, no immediate recall shift appeared in the relief group. To determine, on the other hand, whether a delayed shift occurred for this group or for other groups; namely, for stress Ss who were subsequently relieved, and for nonrelieved stress Ss (the last as a function of time lapse alone), we turn to the data of Table 3 (Columns 1-6). None of the recall differences recorded there (for CR-IR, CR alone, or IR

i.e., which particular tasks were completed and which were incompleting, was found to be a nonsignificant source, and hence its variance was added to the within-cells variance.

<sup>8</sup> These results are similar to those of Glixman (9). Whereas Glixman, however, regards such findings as indicative of a repression process, we are suggesting that they do not point unequivocally to such a conclusion.



TABLE 2

ANALYSES OF VARIANCE OF INITIAL DIGIT RECALL (D), SCRAMBLED SENTENCE RECALL SCORES, AND PERCEPTUAL RECOGNITION SCORES FOR S-RECALLERS (PRIVATE SCHOOL Ss) AND F-RECALLERS (PUBLIC SCHOOL Ss) UNDER THREE EXPERIMENTAL CONDITIONS

Source	D			CR-IR		TR		CR		IR		P	
	df	ms	F	ms	F	ms	F	ms	F	ms	F	ms	F
Type of Ss	1	3.86		0.76		5.80		0.56		2.62	1.19	111.34	2.15
Conditions	2	5.78	1.16	0.96		23.25	3.82*	7.46	3.50*	4.84	2.20	13.12	
Ss × Conditions	2	6.32	1.27	13.57	5.19*	5.50		5.38	2.53	4.55	2.07	458.50	8.84*
Error	110	4.99		2.61		6.09		2.13		2.20		51.87	

\* For 2 and 110 df, an  $F$  of 3.09 is significant at the .05 level; an  $F$  of 4.81 is significant at the .01 level.

TABLE 3

MEAN DIFFERENCE BETWEEN FIRST AND SECOND RECALL SESSIONS, AND BETWEEN FIRST AND THIRD RECALL SESSIONS, FOR CR-IR, CR, AND IR

Condition of Initial Recall	Condition of Delayed Recall	S-recallers (Private School Ss)						F-recallers (Public School Ss)					
		2nd-1st sessions			3rd-1st sessions			2nd-1st sessions			3rd-1st sessions		
		CR-IR	CR	IR	CR-IR	CR	IR	CR-IR	CR	IR	CR-IR	CR	IR
stress	stress	-0.1 ( $N = 13$ )	0.2	0.3	-0.5 ( $N = 12$ )	0.0	0.5	-0.7 ( $N = 10$ )	0.2	0.9	0.1 ( $N = 10$ )	0.2	0.1
stress	relief	0.0 ( $N = 6$ )	0.0	0.0	-1.2 ( $N = 6$ )	-1.0	0.2	0.0 ( $N = 13$ )	0.3	0.3	0.4 ( $N = 13$ )	0.3	-0.1
relief	relief	0.1 ( $N = 19$ )	0.3	0.2	-0.2 ( $N = 16$ )	-0.2	0.0	-0.3 ( $N = 23$ )	0.0	0.3	-0.3 ( $N = 23$ )	-0.2	0.1
neutral	neutral	0.2 ( $N = 15$ )	0.4	0.2	0.2 ( $N = 15$ )	0.1	-0.1	-0.5 ( $N = 18$ )	0.0	0.5	-0.6 ( $N = 18$ )	-0.2	0.4

alone) are significantly different from zero,<sup>9</sup> hence implying no delayed restoration of forgotten items. The present findings thus do not support predictions derived from a repression interpretation of the S-recall pattern, but are consistent with those derived from a selective learning interpretation.

2. *Is the F-recall pattern a function of selective remembering or selective learning in favor of failures?* To answer this question, we must examine the degree of shift in the recall difference score from stress to relief conditions for public school Ss (Table 1, Column 2). No such shift occurred, the two means being identical. Recall of incompleting tasks alone, moreover, shows a nonsignificant increase, whereas the task-tension theory would predict, if anything, a decrease. Recall of completed tasks and total

recall also show negligible increases. Here again, then, no immediate recall shift occurred in the relief condition. Further, there was no delayed shift for any of the subgroups. As indicated in the last six columns of Table 3, none of the differences for CR-IR, CR alone, or IR alone, were significantly different from zero. Our findings, therefore, do not confirm predictions derived from a task-tension interpretation of the F-recall pattern, but once more are consistent with those implied by a selective learning view.

3. *Is there a mnemonic reaction to the test as a whole?* The data relevant to this question are the threshold shifts from stress to relief conditions for test-related words (compared to matched neutral words). According to Table 1, Column 6, private school Ss in the relief group show a significant return ( $p < .05$ , two-tail test) to the level obtained under neutral conditions, whereas public school Ss do not shift significantly. That the perceptual

<sup>9</sup> As can be seen from Columns 5 and 6, the CR-IR difference of -1.2 ( $p < .075$ ) is a function of decreased recall of completed tasks rather than increased recall of incompleting tasks.

defense pattern dissipated for private school Ss when they were relieved, suggests that a genuinely repressive reaction—a memory rather than a learning phenomenon—occurred for these Ss with regard to the threatening situation as a whole. Since the perceptual vigilance reaction of the public school Ss, on the other hand, did not dissipate in the relief condition, it would seem that this reaction is due to a learning rather than a memorial process.

#### DISCUSSION

Three propositions summarize the present results. 1. *Both the S- and F-recall tendencies in the present study are due to a selective learning rather than a selective remembering mechanism.* The pertinent datum here is the following: Among the boys from private schools, i.e., those exhibiting an S-recall pattern, the tendency to recall failures was no greater for relieved than for nonrelieved Ss; likewise, among the boys from public schools, i.e., those exhibiting an F-recall pattern, the tendency to recall failures was no less for relieved than for nonrelieved Ss. 2. *A repression mechanism was found for S-recallers, but with respect to the stress situation as a whole rather than for specifically failed items.* The evidence here: Nonrelieved private school boys manifested perceptual defense for words related in a general way to the examination situation, whereas those who were relieved showed no such perceptual defense. 3. *Instead of a comparable process of enhanced retention of the total stress situation, F-recallers showed enhanced registration for the total situation.* Thus, although public school boys under stress were perceptually vigilant for test-related words (relative to public school boys under neutral conditions), this effect did not dissipate in the relief condition.

Before we can accept these statements, two critical procedural issues require reexamination: namely, (a) was the stress situation sufficiently threatening; and (b) if so, was the relief treatment really therapeutic? Consider first the question of threat.

Needless to say, the present data would have no bearing on a theory of repression if the stress situation had not induced intense ego threat. That such threat did indeed occur may be inferred from several lines of evidence—

observational, introspective, and experimental. In the first place, Ss in the stress condition showed many overt signs of emotional disturbance (little or none of which appeared among the neutral Ss). When stumped by a problem, they would squirm nervously or shake their heads in disgust (sometimes swearing simultaneously) or would stare at the page with a very pained expression. Further, there were many attempts to conceal one's difficulty from others, either by pretending to have finished problems not actually completed, or, when able to solve a problem, by trying to make one's success very obvious to competitors (e.g., putting one's pencil down with a bang or yawning loudly while stretching). Again, depressive and hostile reactions were quite prevalent following the exam, many Ss sitting morosely in their chairs, others openly challenging the worth of the test. Lastly, expressions of great relief—deep sighs, postural relaxation, laughter—almost always appeared when the ruse was finally explained. As to introspective evidence: First, most Ss voluntarily reported that they had been extremely concerned prior to, and tremendously relieved following, exposure of the hoax. They attributed emotions to themselves ranging all the way from intense feelings of inadequacy (“I said to myself, ‘See, you really don’t belong at Harvard’”) to violent hostility (“I wanted to get that supercilious bastard sitting beside me.”) Secondly, continued questioning revealed that although some thought the test a poor one, no one had doubted its reality. With regard to experimental data, finally, consider the results for our indicator of anxiety: digit recall. Whereas recall of digits showed a marked pretest to posttest improvement for Ss in the neutral condition, Ss under stress showed no such improvement, the *F* between conditions being significant well beyond the .01 level. The most plausible explanation for the latter Ss’ deficiency would seem to be the disrupting effects of anxiety generated during the scrambled sentence test. These data, in sum, would seem to support the assumption that intense ego threat was aroused in the present study.

Let us turn now to the second question: Can exposure of the hoax really be considered a therapeutic, i.e., anxiety-reducing, technique?



Even if it were conceded that anxiety had been aroused in our stress condition, the present findings could be legitimately challenged on the ground that this anxiety was not effectively reduced in the relief condition. Such a challenge might proceed as follows: In the typical clinical setting, the therapist's efforts to uncover repressed events meet with strong defensive resistance. Hence, recovery of repressed material requires a great deal of time and involves a special kind of intimacy between patient and therapist, neither of which occurred in the present relief treatment. The point is well taken, but it overlooks two important differences between repression as it ordinarily occurs and repression as precipitated in the present experiment. First, in ordinary circumstances, the fearful event is unknown to the therapist, having occurred before therapist and patient began their relationship. Thus, as long as the therapist is ignorant of the actual situation and cannot confront the patient with reality, the patient may safely continue to hide behind his defense and deny the event. In the present situation, however, the individual is faced with an omniscient therapist—one who produced the threatening event and who is as much an authority on its nature as the patient. To continue repressing in the face of such omniscience is to render one's contact with reality precarious at best. Second, and perhaps more critical, there would seem to be but two ways in which a threatening event may be made less fearful: (a) by effecting a radical change in the patient's motivational structure (e.g., eliminating his fear of intellectual incompetence) so that the event (e.g., exam failure) no longer poses a threat; or (b) by altering the significance of the event so that it is no longer motivationally relevant. The former is the proper goal of the therapist in the usual clinical setting. The latter is what was obviously involved in the revelation of the experimental hoax to our Ss: the failure experience had initially been created by deception and was now being restored to its true meaning as a psychological gimmick. E's revelation does not make S any less of a repressor or eliminate his problem with respect to intellectual competence (nor was it intended to), but merely renders this particular event irrelevant to that problem and hence no

longer something to be defended against. In the light of both these points, then, the assumption that effective therapy occurred in the present experiment seems entirely tenable.

Two general conclusions follow from the present findings. One is methodological, and touches on the question of criteria for a suitable test of repression. The other concerns the particular mechanisms that underlie the avoidant and adient cognitive reactions obtained here. Consider each of these in turn.

1. In attempting to provide an experimental demonstration of repression, the general practice has been to compare the recall of certain "nonthreatening" aspects of the experimental situation with that for certain threatening aspects. In particular, it has been common to examine the recall of failures in contrast to successes on a bogus intelligence test. The present results impute this approach, for they suggest that repression embraces only the most general aspects of the stress situation, whereas recall of particular threatening or "nonthreatening" aspects of that situation is a function of learning. This finding seems entirely reasonable, for when a person has failed half the items on a test, what is likely to be threatening for him is not that he has failed those particular items but that he has failed the *test*. Nor can the items he has passed be entirely free from threat if the total situation in which they are embedded arouses anxiety. It would thus seem necessary, in experimental research on repression, to determine precisely what category of materials is relevant to the subject's conflicts and fears. While recent experiments in this area have been careful to satisfy the previously neglected criterion of motivational relevance of experimental materials, the present findings strongly suggest it would also be well to consider the cognitive relevance of these materials.

2. That both public and private school Ss showed opposite and significant differences between stress and neutral conditions, and that the directions of these differences were completely consistent across recall and perception, indicate that two quite different defensive reactions to stress occurred in this study. One—an avoidant reaction by private school Ss—was indeed shown to be repressive as far as the total test situation was concerned, but, with regard to particular items, was a

matter of registration differences. The other—an adient reaction by public school Ss—was a case of registration differences at both levels. One may properly inquire at this point what the particular processes might be which determine the various registration differences found. Osgood (15), among others, has suggested that differential rehearsal may account for such learning superiorities. The possibility of covert review of failures by Harvard public school Ss indeed seems especially reasonable on the assumption that these Ss are high achievers. While such individuals are not unconcerned about failure, they undoubtedly have a history of repeated reward for efforts to counteract initial failure. Hence these Ss would be expected to ruminate over their difficulties as a cognitive precursor to constructive adjustment. In the case of private school Ss, covert review of successes might be expected as a kind of ego-bolstering reaction. Another possibility for these Ss is that anxiety over inability to solve items acts to disrupt their registration. These two mechanisms are, of course, by no means mutually exclusive.

#### SUMMARY

The present study proposed to answer three questions: 1. Is superior recall of successes in an intelligence test situation (S-recall) a function of selective forgetting (repression) of failures or selective learning in favor of successes? 2. Is superior recall of failures in this situation (F-recall) a function of selective remembering or selective learning in favor of failures? 3. Is there a mnemonic reaction to the test as a whole? The first two questions were examined by comparing the recall preference of nonrelieved S- and F-recallers on a bogus intelligence test (involving solvable and unsolvable scrambled sentences) with that for comparable relieved Ss, selective remembering theories predicting a shift from nonrelieved to relieved conditions for each recall group. The third question was examined by comparing the perceptual recognition thresholds of the above relieved and nonrelieved Ss on exam-related and matched neutral words. Forty-two Ss each were tested in the relief and nonrelief condition, with 33 more tested under neutral conditions. Prior attendance at a private

versus a public high school was used as the indicator of S- and F-recallers respectively.

The findings were as follows: 1. Both the S- and F-recall tendencies were due to a selective learning rather than a selective remembering mechanism. 2. A repression mechanism was demonstrated for S-recallers with regard to the stress situation as a whole. 3. A comparable process of enhanced retention of the total stress situation was not demonstrated for F-recallers; rather, the effect here was due to enhanced registration. Two conclusions were drawn from these findings: First, that repression was found for the total test situation rather than for specific items, suggests experimental tests of repression must fulfill a criterion of cognitive relevance of test materials. Second, the kinds of enhanced registration found may have been due to covert rehearsal, and possibly also to disruption of the registration of other items by anxiety.

#### REFERENCES

1. ALPER, THELMA G. Memory for completed and incomplete tasks as a function of personality: An analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403-420.
2. ALPER, THELMA G. Memory for completed and incomplete tasks as a function of personality: correlation between experimental and personality data. *J. Pers.*, 1948, **17**, 104-137.
3. ATKINSON, J. S. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381-390.
4. BLUM, G. S. *Psychoanalytic theories of personality*. New York: McGraw-Hill, 1953.
5. CARON, A. J. The effects of task, objective orientation, and personality factors on recall in an interrupted-task situation. Unpublished doctoral dissertation, Univer. of Chicago, 1954.
6. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230-235.
7. ERIKSEN, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952, **44**, 442-447.
8. ERIKSEN, C. W. Psychological defenses and "ego strength" in the recall of completed and incomplete tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45-50.
9. GLIXMAN, A. F. Recall of completed and incomplete activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281-295.
10. GUILFORD, J. P. *An Inventory of Factors STDCR*. Beverly Hills, Cal.: Sheridan Supply, 1940.
11. KNUTSON, A. L. Personal security as related to station in life. *Psychol. Monogr.*, 1952, **66**, No. 336.
12. LEWIN, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.



13. MCARTHUR, C. Personalities of public and private school boys. *Harvard Educ. Rev.*, 1954, **24**, 256-262.
14. MCARTHUR, C. Personality differences between middle and upper classes. *J. abnorm., soc. Psychol.*, 1955, **50**, 247-254.
15. OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.
16. RAPAPORT, D. *et al. Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.
17. ROSENZWEIG, S. An experimental study of "repression" with special reference to need-persistent and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64-74.
18. ROSENZWEIG, S., & SARASON, S. An experimental study of the triadic hypothesis: reaction to frustration, ego defense, and hypnotizability. I. Correlational approach. *Charact. & Pers.*, 1942, **11**, 1-20.
19. SCODEL, A. Passivity in a class of peptic ulcer patients. *Psychol. Monogr.*, 1953, **67**, 1-15.
20. UNDERWOOD, B. J. Learning and retention. In C. P. Stone (Ed.), *Annual review of psychology*, Vol. III. Stanford: Annual Reviews, 1953.
21. WECHSLER, D. *The measurement of adult intelligence*. (3rd ed.) Baltimore: Williams & Wilkins, 1944.
22. ZEIGARNIK, BLUMA. Ueber das Behalten von erledigten und unerledigten Handlungen. *Psychol. Forsch.*, 1927, **9**, 1-85.
23. ZELLER, A. F. An experimental analogue of repression. I. Historical summary. *Psychol. Bull.*, 1950, **47**, 39-51.

Received June 17, 1957. Prior Publication.

# GROUP CLIMATE AS A FACTOR IN THE RELEVANCE AND ACCURACY OF SOCIAL PERCEPTION<sup>1</sup>

RALPH V. EXLINE

*Fels Group Dynamics Center, University of Delaware*

**T**HIS PAPER reports on a study of accuracy of social perception as a function of the "climate" of social interaction. Special attention is paid to the relevance of the perceived, estimated, or judged social behavior of others to the goals of the perceiver.

Much of the research on accuracy of social perception has been concerned with accuracy as an attribute of the individual perceiver (3, 8, 12). Results of certain of these studies, however, suggest the merit of studying the relationship between accuracy of perception and the social context in which perception is measured. Conrad (5) and Chowdhry and Newcomb (4), for example, suggest that accuracy is a function of the relevance of the perceptual task to the motives of the perceiver, while Suchman (22) speculated that group differences in accuracy of task-oriented perception are caused by the nature of the interaction within the groups studied.

The theoretical relationship often assumed to exist between motivation and perception suggests that the study of perceptual accuracy in a social context may be facilitated by considering perception in terms of a path-goal relationship. Thus, according to Lewinian theory (16), once a person defines a goal, he attempts to move from a position outside to a position within the goal region. In a relatively unstructured situation where a goal, but not the means to achieve it, has been defined, movement toward the goal requires that he direct energy to the differentiation of a path leading toward the goal. This thinking can be generalized to group situations and seems especially pertinent where information concerning the opinion of other group members is a necessary part of the path region (and where the opinion of others can be learned, or inferred, from their behavior). In these cases, persons can be expected to direct energy to the understanding of relevant opinions. Hence, every-

thing else being equal, they should be more likely to be aware of and to perceive accurately those behaviors of others which they define as relevant to locomotion toward the goal.

In the light of research and theory (7, 16, 18) that conceives of the group as providing a psychological social environment for its members, the environment created by the group may have predictable effects on the likelihood that a given behavior is seen as a "path" element and hence on the accuracy with which it is perceived. It is the general hypothesis of this study that the group environment, through its effect upon what is perceived as relevant to the goals of the group members, predictably affects the accuracy with which specified social phenomena are perceived.

One aspect of the psychological social environment that groups provide their members is that of the social "climate" (18) or "hedonic tone" (15) that characterizes the interaction of group members. The relationship of this phenomenon to goal-striving activity of individuals has been widely documented. Research reported by Lippitt and White (18), Deutsch (7), and Hemphill and Westie (15) has shown that friendliness and congeniality of the group's "social climate" occurs concomitantly with motivation to persist in working toward task goals assigned to the group. There is also evidence to indicate that individuals as members of groups often pursue individual, or ego-centered goals, rather than ones oriented to the group task. The social climate of these latter groups has been shown to be one of less pleasantness (13), more conflict (10), and less friendliness (7) than the climate of groups where members' goal-orientations are less individual in nature.

The effect of the group's climate upon the relevance of a given goal, hence upon the nature of behavior seen as defining the goal path, provides the bridge between the group environment and the accuracy with which individuals perceive specified behaviors. Thus, if members of congenial groups are more likely

<sup>1</sup> This report is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Illinois.



to hold a group goal in common, opinions that bear on the group goal are more likely to be mutually relevant, hence more accurately perceived, than in groups of less congenial atmosphere. Conversely, if members of groups characterized by uncongenial interaction climates are more prone to hold individual, ego-centered goals, then hypothetically, opinions held and communicated by others that are relevant to one's own ego-centered goals should be more relevant, hence more accurately perceived.

On the assumption that task-oriented activity is perceived to be more relevant in congenial than in noncongenial group climates, while ego-oriented activity is perceived to be more relevant in noncongenial than in congenial group climates, the discussion above may be more formally stated in terms of two hypotheses which this study is designed to test.

*Hypothesis I: Members of congenial groups are more accurate than members of noncongenial groups in perceiving task-oriented behaviors.*

*Hypothesis II: Members of noncongenial groups are more accurate than members of congenial groups in perceiving ego-oriented behavior.*

## METHOD

### Subjects

One hundred thirty graduate students assumed to be relatively homogeneous as to age, intelligence, and socioeconomic status (they were public school teachers) were recruited from a summer session at the University of Illinois. To minimize the effects of previous acquaintance, each group was composed of Ss drawn from a different section of a basic course in educational psychology. The fact that these students were teachers whose regular jobs were in communities scattered throughout the state was taken to be an additional restriction on acquaintance.

### Design

The Ss were randomly assigned to two treatments by group atmosphere, then subdivided into groups of five people of the same sex. The groups were first given instructions designed to create a congenial or noncongenial atmosphere. Following these instructions, each group worked on a standard discussion task, after which the members marked a set of rating scales designed to measure the accuracy of their perception of task-oriented and ego-oriented behaviors of others, their satisfaction with the progress of the group, and their willingness to continue working with the group.

## Operational Definitions<sup>2</sup>

*Congenial groups.*<sup>3</sup> Congenial groups were those groups which, by a process of random selection, were given information designed to create the expectation that group members would be congenial to one another. At the time of recruitment, Ss were given a standard personality test and a questionnaire purported to measure attitudes towards working in groups, and told that this information would be used to help the investigator assign them to discussion groups in which they would find the other members congenial. Upon arriving at the laboratory, Ss assigned to the congenial treatment were told that they stood very high on a scale of congeniality developed from the preliminary tests—"the matching was so good as to be expected only once or twice in all the groups studied," and that they should get on well together.

*Noncongenial groups.* These groups were given information designed to create the expectation that group members would not be congenial to one another. The Ss assigned to these groups were told that due to the limited number of available subjects, not all groups could be matched equally well as to congeniality. They were told that they stood very low on the congeniality scale, that it was necessary to put people with opposing views in the same group, that they were about as bad a matching as was possible, and that they probably would not find their group very congenial.

*Task.* The task was introduced as one that would enable psychologists to relate the Ss' scores on the previous tests to their ability to take the role of another person. It was implied that psychologists think this ability important in establishing good human relationships with other people.

The Ss were asked to role-play a group of trainees in vocational guidance, meeting to discuss the disposition of a vocational counselee. They were to reach agreement upon a recommendation concerning the disposition of the case. Each S was given the identical test and biographical information about the hypothetical client, but in addition received a unique recommendation based on one of five different orientations toward the data. Thus, one S suggested a job on the basis of an orientation that emphasized the need to consider present and future job opportunities. Another S's suggestion was based on the relationship between personality factors and job requirements, a third on the role of training and past experience, a fourth on test scores, and the fifth on test scores in combination with training and past experience. The disagreement that was thus ensured was expected to facilitate the production of behavior indicating interpersonal preference as well as of differential orientation toward the path to task completion.

A standard amount of time was allowed for Ss to

<sup>2</sup> Complete descriptions of the instructions, results by individual Ss and detailed statistical tables can be obtained in reference (9) from University Microfilms, Ann Arbor, Michigan.

<sup>3</sup> Instructions designed to create congenial and noncongenial groups were a modified version of those used by Back (1).

familiarize themselves with the data and the point of view they were to support in the discussion. To insure that each group had a minimum opportunity to hear all points of view, the instructions required each person to present his suggestion and opinion in turn. After each *S* had spoken twice, a general discussion was permitted.

**Task-oriented behavior.** Task-oriented behavior is defined conceptually as the manipulation, symbolic or otherwise, of the idiosyncratic content of the task. Its operational definition, like that of ego-oriented behavior, becomes complex when the individual is a member of a problem-solving group, for the definition must involve reaction to others' acts. Action and reaction are necessary to the processes of communication, evaluation, control, decision, and tension reduction which according to Bales (2) characterize the sequence of group problem solving. Individuals in such groups communicate not only their own orientations to the problem, but their evaluations of other's orientations. Both such types of communications qualify as task-oriented behaviors in that both are necessary to the movement of the group toward its goal of decision. On the basis of these considerations, task-oriented behavior is operationally defined for present purposes as verbal or nonverbal behavior that communicates either the frame of reference used by a person to organize the task data, or his reaction to the frames of reference used by others to organize the same data.

**Ego-oriented behavior.** The individual is said to be ego oriented when his goal is the enhancement of self without regard to the demands of the task environment (17). Ego-oriented behavior refers to the manipulation of elements in the task situation that bear on the self as a value. It is assumed that in the process of working through disagreements, *Ss* communicate information pertinent to their orientation to and evaluation of their fellow group members as persons. Ego-oriented behavior, then, is operationally defined as verbal or nonverbal behavior that communicates a person's relative acceptance of other group members.

**Accuracy of perception of task-oriented behavior.** Task-oriented perceptual accuracy is defined as the discrepancy between each *S's* rank ordering of the use of four orientations toward counseling as he thinks the other *Ss* would rank order them, and the other *S's* actual rank ordering of the same four orientations.

An error score for each *S* was obtained as follows. Each *S* was given a booklet containing the definitions of the orientations toward counseling and five sets of rating scales. Each scale consisted of 20 intervals distributed over six definition points from "never used" to "exclusive use." Each *S* filled out one set of ratings to indicate the extent to which he felt he had used the four independent orientations<sup>4</sup> in his own discussion. Each *S* also filled out the remaining sets of scales to indicate his perception of the rank order of the extent to which each of the others favored the various orientations. Thus, each *S's* own responses provided a key against which the estimates made by the others could be scored for errors. An over-all error score representing *S's* ability to accurately predict the rank order assigned

to each orientation by each *S* was obtained by averaging *S's* squared errors over all orientations and others.

Accuracy scores defined in terms of a rank-order metric have the advantage of eliminating the possibility that low error scores merely reflect the judges' willingness to use extremes, or ability to judge deviations of others from a central tendency. Scores derived by this method reflect a component of accuracy that Cronbach (6) terms *SAr*, i.e., accuracy in judging the rank ordering of the items made by the group as a "generalized other." Since items correspond to task orientations, this type of accuracy is assumed to be relevant to the problem studied.

**Accuracy of perception of ego-oriented behavior.** Ego-oriented accuracy was measured in a manner similar to that of task-oriented accuracy. In this case, the *S* rated the extent to which he liked the others and estimated the ratings each of the others gave their co-workers. Ratings were made on scales which, but for re-worded designation points, were identical with those used for task-orientation. An error score was derived as described in task-oriented perception. In this case, however, the other members of the group were the items which *Ss* rank-ordered.

## RESULTS

### *Effectiveness of Manipulations*

To establish that the treatments did indeed create the necessary differences in congeniality, the extent to which members of the various groups liked their co-workers, and their willingness to continue with the group, were assessed. Members of "congenial" groups marked a scale of liking of co-workers significantly higher than did the members of "noncongenial" groups (means of 16.05, 14.60; *SDs*, 1.99, 1.95;  $p = .02$ , two tails). Similarly, the "congenials" rated themselves significantly higher ( $p = .02$ ) in willingness to continue working with the group (means of 16.72, 15.19; *SDs*, 3.26, 3.92). Since *Ss* assigned to congenial groups showed no greater pre-experimental disposition to work with groups than did *Ss* assigned to noncongenial groups<sup>5</sup>, the above data are taken as evidence that the inductions created the desired climates of congeniality.

<sup>5</sup> The conclusion that *Ss* in congenial and noncongenial groups did not differ in their pre-experimental disposition to work in groups, was based on the response to an item in the questionnaire, given to *Ss* in their classes, which measured attitudes toward working in groups. A low score on this item signified greater willingness to work in groups. Means of *Ss* in congenial and noncongenial groups were, respectively, 2.41 ( $N = 65$ ,  $SD = .94$ ) and 2.35 ( $N = 59$ ,  $SD = .92$ ). Six *Ss* in noncongenial groups did not answer the question. A  $t$  of .97 significant at greater than the .30 level (two-tailed test) was obtained.

<sup>4</sup> The orientation of the fifth *S* was a combination of those of the third and fourth *Ss*.



TABLE 1

MEAN ERROR SCORE IN ESTIMATING TASK- AND EGO-ORIENTED BEHAVIOR MADE BY SUBJECTS CATEGORIZED AS TO CONGENIALITY TREATMENTS AND SEX

Perception Measure	Treatment			
	Congenial		Noncongenial	
	N	$\bar{X}$	N	$\bar{X}$
Task-oriented				
Male subjects	32	7.16	32	8.06
Female subjects	32	6.51	32	7.89
Ego-oriented <sup>a</sup>				
Male subjects	20	9.55	25	10.11
Female subjects	30	7.98	19 <sup>b</sup>	8.96

<sup>a</sup> Mean-per-person error is based on five-person groups only. The difficulty of this estimation was not comparable in the five- and four-person groups.

<sup>b</sup> One S filled out scales improperly.

### Tests of Hypotheses

The hypotheses concerning the relationship between social climate and the accuracy of perception of task and ego-oriented behavior of others were tested by applying analysis of variance techniques to mean-per-person error scores. Although the problem was not formulated in terms of sex differences, the composition of the groups permits this comparison.

**Hypothesis I.** Mean-per-person error scores on task-oriented perception for congenial and noncongenial groups are listed in Table 1 and the results of the analysis of variance in Table 2. The results indicate that members of congenial groups, regardless of sex, are significantly more accurate than members of noncongenial groups in task-oriented perception.

**Hypothesis II.** Mean-per-person error scores on ego-oriented perception for congenial and noncongenial groups are listed in Table 1 and the results of the analysis of variance in Table 3. An examination of ego-oriented perception error scores shows no statistically significant differences between congenial and noncongenial groups, although significant differences attributable to sex are readily apparent.

Hypothesis I is thus confirmed by the data, while Hypothesis II is not. Indeed, the data indicated a trend, albeit a statistically insignificant one, in the direction of greater ego-oriented accuracy on the part of members of congenial groups. An unanticipated but interesting finding is that women are more accurate than men in the estimation of their co-workers' interpersonal likes and dislikes.

TABLE 2

ANALYSIS OF VARIANCE OF TASK-ORIENTED ESTIMATION ERROR OF SUBJECTS CATEGORIZED BY SEX AND CONGENIALITY TREATMENT

Source of Variation	df	Mean Square	F
Sex	1	5.36	.85
Congeniality treatment	1	41.63	6.62*
Interaction: sex X congeniality	1	1.95	.31
Within groups	124	6.29	

\* Significant at .05 level or beyond.

TABLE 3

ANALYSIS OF VARIANCE OF EGO-ORIENTED ESTIMATION ERROR OF SUBJECTS CATEGORIZED BY SEX AND CONGENIALITY TREATMENT

Source of Variation	df <sup>a</sup>	Mean Square	F
Sex	1	43.47	4.90*
Congeniality treatment	1	14.33	1.62
Interaction: sex X treatment	1	.71	.08
Within groups	90	8.87	

\* Significant at the .05 level or beyond.

<sup>a</sup> Tsao's (24) approximation method was applied to correct for disproportionate subclass n.

### Post Hoc Analyses

Examination of Ss' responses to the post-meeting questionnaire provide additional data concerning the relationship between group climate and accuracy in perceiving task and ego-oriented behavior. It has already been mentioned that Ss were asked to indicate their willingness to continue working with the group on similar tasks in order to measure the effect of the congeniality inductions. If this measure is a valid index of congeniality, Ss more willing to continue should perceive task-oriented behavior more accurately than do those who find the group less attractive, regardless of experimental treatment; while, conversely, those less willing to continue should perceive ego-oriented material more accurately.

These *post hoc* hypotheses were tested by first categorizing Ss according to whether their willingness to work with the same group was above or below the median of the pooled scores, then applying analysis of variance techniques to determine the effects of congeniality inductions, willingness, and the interaction between the two upon the social perception error scores. Mean-per-person error scores are listed in

TABLE 4  
MEAN ERROR SCORE IN ESTIMATING TASK- AND  
EGO-ORIENTED BEHAVIOR MADE BY SUBJECTS  
CATEGORIZED AS TO WILLINGNESS TO  
CONTINUE AND CONGENIALITY

Perception Measure	Treatment			
	Congenial		Noncongenial	
	N	$\bar{X}$	N	$\bar{X}$
Task-oriented				
High willingness	37	6.67	27	7.13
Low willingness	27	7.06	37	8.58
Ego-oriented				
High willingness	29	9.45	18	10.04
Low willingness	21	7.89	26	8.97

TABLE 5  
ANALYSIS OF VARIANCE OF MEAN ERROR SCORES IN  
TASK-ORIENTED PERCEPTION MADE BY MORE  
AND LESS WILLING SUBJECTS

Source of Variation	df <sup>a</sup>	Mean Square	F
Willingness	1	27.08	4.593*
Congeniality treatment	1	31.36	5.318*
Interaction: willingness $\times$ congeniality	1	8.99	1.524
Within groups	124	6.12	

\* Significant at the .05 level or beyond.

<sup>a</sup> Tsao's (24) approximation method was used to correct disproportionate subclass n.

Table 4 and the results of the analysis of variance in Tables 5 and 6.

Inspection of the means in Table 4 shows that, in both congenial and noncongenial groups, Ss who were more willing to continue with the group made fewer task-oriented errors than did their less willing co-workers. It is also apparent that in both congenial and noncongenial groups, fewer errors in assessing interpersonal relations were made by those less willing to continue. Since the *F* ratios in Tables 5 and 6 indicate that effects of willingness are significant at better than the .05 level, the data support the hypotheses.<sup>6</sup>

<sup>6</sup> Although no data concerning the productivity of the groups were collected, it is of interest to note that members of congenial groups expressed greater satisfaction with the progress of their groups. Mean satisfaction ratings, on a 20-interval scale similar to those described previously, were noted as follows: Congenial male and female, 14.44 and 15.47 respectively; noncongenial male and female, 12.19 and 11.06 respectively. With mean squares of 354.44, .07, 37.19, and 18.17 for congeniality, sex, interaction, and error variance, respectively, the only significant *F* was that of 19.51 (significant at the .001 level with *df*'s of 1 and 124)

TABLE 6  
ANALYSIS OF VARIANCE OF MEAN ERROR SCORE IN  
EGO-ORIENTED PERCEPTION MADE BY MORE  
AND LESS WILLING SUBJECTS

Source of Variation	df <sup>a</sup>	Mean Square	F
Willingness	1	40.64	4.089*
Congeniality treatment	1	16.38	1.684
Interaction: willingness $\times$ congeniality	1	1.41	.014
Within groups	90	9.94	

\* Significant at the .05 level or beyond.

<sup>a</sup> Tsao's (24) approximation method was used to correct disproportionate subclass n.

## DISCUSSION

Results reported above call into question the generality of a theory of accuracy of social perception that postulates accuracy as a function of the relevance of the perceived behavior to the goals of the perceivers. On the one hand, the finding that task-oriented phenomena are more accurately perceived in congenial group atmospheres, where they are assumed to be more relevant to the goals of group members, would appear to provide support for the theory. On the other hand, the theory was not supported by the finding that ego-oriented phenomena are not more accurately perceived in the noncongenial group condition, the setting where ego-orientation was assumed to be more relevant.

It is possible that the operational measure of ego-oriented accuracy of perception did not provide a valid test of the theory with reference to such behavior. There is evidence to suggest (1, 2, 13) that the mapping of interpersonal relations is a relevant activity for members of congenial groups.

On the other hand, existing theory and data suggest that regardless of the validity of operationally defining ego-relevant activity as the assessment of interpersonal relations, one could not expect such assessment to be accurately carried out. Newcomb's (19) theory of communicative acts, for example, suggests that accuracy of perception of others' behavior in a group is a function of the extent to which

found for congeniality. It is uncertain whether the difference noted above is related to the greater task-oriented accuracy recorded in congenial groups or merely reflects the impact of the congeniality inductions. In view of this uncertainty these scores are not discussed further.



the behavior can be said to be an object common to group members. Thus, while it is possible to conceive of task-oriented behavior as an object common to the members of congenial groups, it is less easy to see how others' behavior perceived as relevant to the private, ego-oriented goals of one group member can be said to be held in common by all group members. Also relevant to this discussion are data reported by Tagiuri, Blake, and Bruner (23), who indicated Ss could not accurately perceive cues of rejection. To the extent that such cues are more likely to occur in noncongenial than in congenial groups, a factor favoring greater accuracy of perception of interpersonal relations exists in congenial groups.

In spite of the above arguments, there remains the fact that interpersonal relations were perceived with greater accuracy by those less willing to continue their association with the group. It was suggested, moreover, that this result could be interpreted as supporting the hypothesized relationship between the congeniality of the group and accuracy of perception of ego-relevant behavior.

An alternate interpretation may explain this relationship. This interpretation, which assumes that interpersonal conflict occurs in the groups, is at variance with the specific derivation embodied in Hypothesis II, but not necessarily with the general theoretical position of this paper (conflict, after all, is a feature of the group's environment). If cues concerning interpersonal relations are more likely to be observed in situations of interpersonal conflict, and if one can assume that conflict adversely affects a member's willingness to continue with the group, it follows that accuracy in perception of interpersonal relations and unwillingness to continue with the group may be concomitant results of conflict. Unfortunately, there are no measures of the extent to which such conflict marked the process of group discussion. The question of the meaning of more accurate perception of interpersonal relations on the part of those less willing to continue with the group thus remains unanswered. In view of the speculative, even *post hoc*, nature of the postulated relationships among willingness, congeniality, and accuracy of perception, it is suggested that generalization of the theory herein presented be limited to the perception of task-oriented phenomena. The

significant relationships noted between willingness and the two kinds of accuracy of perception, however, do suggest the desirability of further exploring the nature of these relationships, as well as their implications for the processes of problem-solving groups.

Finally, the unanticipated finding that women perceive interpersonal relations more accurately than men appears to merit some discussion. This difference is not related to sex differences in willingness to work with the group, a factor already shown to be inversely related to such accuracy, for no differences between the willingness of men and women were found in either treatment. We must look further for an explanation of this phenomenon.

One explanation of this finding rests on the assumption that our culture sanctions freer expression of feeling by women than by men. It would follow, from this assumption, that more women than men would openly express their likes and dislikes concerning their co-workers. This explanation would appear to be related to the theory of sex-role differentiation developed by Parsons and Bales (20), and studies in jury deliberations by Strodbeck and Mann (21). The results herein reported might be interpreted as further and more general confirmations of the above theory concerning woman's sex role.

An alternative explanation might be phrased in terms of social power. If the general position of women in our society is one which allows them less social power than men, it may be important for them to be able to accurately assess the feelings, attitudes, and relationships among others in order to achieve their personal goals.

Finally, women may be higher in "affiliation motivation" than men. If this is so, French and Chadwick's (11) finding that those having high affiliation motivation are more accurate in perceiving their own popularity status would be pertinent to our data. Research designed to test the validity of the above speculations is in progress.

#### SUMMARY

Twenty-eight groups of five or four graduate students each, controlled for sex, education, and acquaintance were assigned to congenial or noncongenial group climates. Analysis of variance was used to test the effects of group

climates, motivational levels, and the importance of sex difference upon perceptions of task and ego-oriented behavior.

The findings were as follows:

1. Individuals in congenial groups were significantly more accurate in perceiving task-oriented behavior of their group than were members of noncongenial groups.

2. Individuals in noncongenial groups were not more accurate in perceiving the interpersonal relations in the groups than were individuals in congenial groups.

3. Individuals willing to continue to work with the group on a similar task were more accurate in perceiving task-oriented behavior of their group than were individuals less willing to continue with the group.

4. Individuals willing to continue to work with the group on a similar task were significantly less accurate in perceiving interpersonal relations within their group than were individuals less willing to continue with the group.

5. Women, though neither more nor less willing to continue with their group, were significantly more accurate than men in perceiving interpersonal relations.

It is concluded that social climates can predictably affect the perception of task-oriented behavior variables, but that further research on the relationships among social climate, ego-oriented behavior, and accuracy in social perception is necessary. Research into the determinants of sex differences in perception of interpersonal relations is recommended.

#### REFERENCES

1. BACK, K. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9-23.
2. BALES, ROBERT F. *Interaction process analysis*. Addison-Wesley Press, 1950.
3. BIERI, J. Cognitive complexity simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 263-268.
4. CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51-57.
5. CONRAD, H. S. The personal equation in ratings: IIA systematic evaluation. *J. educ. Psychol.*, 1933, **24**, 39-46.
6. CRONBACH, L. J. Processes affecting scores on "Understanding of others" and "Assumed similarity." *Psychol. Bull.*, 1955, **52**, 177-194.
7. DEUTSCH, M. The effects of cooperation and competition upon group process. In D. Cartwright and A. Zander (Eds.), *Group dynamics: Research and theory*. Evanston, Illinois: Row Peterson, 1953. Pp. 319-353.
8. DYMOND, ROSALIND F. A preliminary investigation of the relation of insight and empathy. *J. consult. Psychol.*, 1948, **12**, 228-233.
9. EXLINE, R. V. Relevance as a determiner of the difficulty and validity of social perception. Unpublished doctoral dissertation, Univer. of Illinois, 1954.
10. FOURIEZOS, N. T., HUTT, M. L., & GUETZKOW, H. Self-oriented needs in discussion groups. *J. abnorm. soc. Psychol.*, 1950, **45**, 682-690.
11. FRENCH, E. G., & CHADWICK, I. Some characteristics of affiliation motivation. *J. abnorm. soc. Psychol.*, 1956, **52**, 296-300.
12. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
13. GERARD, H. B. The anchorage of opinions in face-to-face groups. *Hum. Relat.*, 1954, **7**, 313-326.
14. GUETZKOW, H., & GYR, J. An analysis of conflict in decision-making groups. *Hum. Relat.*, 1954, **7**, 367-382.
15. HEMPHILL, J. K., & WESTIE, C. The measurement of group dimensions. *J. Psychol.*, 1950, **29**, 325-342.
16. LEWIN, K. Behavior and development as a function of the total situation. In D. Cartwright (Ed.), *Field theory in social science*. New York: Harper, 1951.
17. LEWIS, HELEN B. An experimental study of the role of the ego in work. *J. exp. Psychol.*, 1944, **34**, 113-126.
18. LIPPITT, R., & WHITE, R. The "social climate" of children's groups. In R. Barker, J. Kounin, & H. F. Wright (Eds.), *Child behavior and development*. New York: McGraw-Hill, 1943.
19. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393-404.
20. PARSONS, T., & BALES, R. F. *Family socialization and interaction process*. Glencoe, Ill.: Free Press, 1955.
21. STRODTBECK, F. L., & MANN, R. D. Sex role differentiation in jury deliberations. *Sociometry*, 1956, **19**, 3-11.
22. SUCHMAN, J. R. Social sensitivity in the small task-oriented group. *J. abnorm. soc. Psychol.*, 1956, **52**, 75-83.
23. TAGIURI, R., BLAKE, R., & BRUNER, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.*, 1953, **48**, 585-592.
24. TSAO, F. General solution of the analysis of variance and covariance in the case of unequal or disproportionate numbers of observations in the subclasses. *Psychometrika*, 1946, **11**, 107-128.

Received March 4, 1957. Prior publication.



# CRITIQUE AND NOTES

## ON THE PHENOMENA OF "OPPOSITE SPEECH"<sup>1</sup>

• BERNARD KAPLAN

Clark University

**I**N A RECENT case study, Laffal *et al.* (16) reported a schizophrenic patient exemplifying "an unusual language syndrome" which the authors refer to as "opposite speech." This syndrome "consists basically in the use of 'yes' by the patient when he means 'no,' and vice versa" (16, p. 409). On a level closer to the data, the syndrome seems to be manifested in the interchange of words of contrasting signification; thus, the patient appeared to interchange "right" and "wrong," "do" and "don't," "like" and "hate," "always" and "never," "any" and "none," etc. Such opposite speech phenomena were manifested in both reception and expression, in both oral and written discourse, and in both spontaneous utterances and responses to questions and test problems. The patient was seemingly unaware of the interchanges; furthermore, there was no analogy to this type of speech activity on the level of concrete, practical action.

In their discussion, Laffal and his collaborators emphasized the different needs of the patient which the opposite speech phenomena might have served; they also considered the inadequacies of contiguity-reinforcement theory in accounting for these phenomena. However, with respect to the nature of the syndrome in the light of cognitive development, the authors limited themselves to a few preliminary remarks and did not take up the issue subsequently. Confronted by what Arieti has described as the schizophrenic's "teleological regression" (2, p. 192), they dealt predominantly with the ostensible teleology and secondarily with the nature of the regression.

It is the purpose of this paper to consider "opposite speech" within the categories of a theory of the development of cognition—a theory concerned with general, formal properties of cognitive activity, obtaining in phylogenesis, ontogenesis, cultural evolution, psychopathology, etc. (24, 25, 26, 27).

The fundamental principle of this comparative-developmental approach to cognition is that wherever development occurs, it proceeds from a relatively global and undifferentiated state to one of

increasing differentiation, articulation, and integration; on the other hand, where development is undone, there is a return to a greater autonomy of local functions and structures. These correlative principles have the status of heuristic laws (cf. 22); though not themselves subject to empirical test, they are valuable in the formulation of empirically testable generalizations.

One such empirical generalization is that adult schizophrenics, in deviating from normal adult modes of organizing experience, move in the direction of a relative dedifferentiation of their cognitively constructed world, and toward a greater autonomy of local functional patterns organized on a lower level of integration (e.g., toward a greater autonomy of concrete, practical activity with respect to linguistically articulated, conceptual thought).<sup>2</sup>

Insofar as cognitive activity in schizophrenics undergoes this dedifferentiation and disintegration, it will show similarities in structure (cf. 23, p. 27 ff.) to cognitive activity occurring under a variety of conditions and among certain groups of nonschizophrenic individuals. Thus, schizophrenic cognitive activity will tend to show formal similarities to the activities of normal man (a) under transitory conditions of stress, anxiety, or distraction; (b) in the process of expressing experiences in primitive media of symbolization, e.g., dreams, gestures, etc.; (c) in earlier stages of sociocultural linguistic codifications of experience. Schizophrenic cognitive activity may also be expected to show similar formal properties to those reflected in the linguistic articulation of experiences among (d) very young children; and (e) such organically dedifferentiated individuals as certain aphasics.

There follows (a) a critical discussion of the findings presented by Laffal *et al.*; (b) an application of the developmental principle to show why "opposite speech" phenomena are regarded as involving a more or less primitive cognitive activity; (c) a summary of the empirical evidence

<sup>2</sup> Dedifferentiation or regression are not here considered as processes of delamination. Regression may be manifested in a return to certain more primitive levels of cognitive activity without entailing the impairment of "schemata" or "apparatuses" structured in the organism in the course of mastering other, genetically more advanced, operations. Furthermore, regression does not here signify a re-evoking of old content, but rather a return to more primitive forms of thought and patterns of operations (cf. 2, 5).

<sup>1</sup> This paper derives, in part, from research and study carried out within a project on normal and pathological symbolization and communication. This project, directed by Heinz Werner and the author, is supported by a grant from The Foundations' Fund for Research in Psychiatry.

from each of the five above-mentioned domains in order to evaluate the thesis that "opposite speech" is more characteristic of relatively primitive cognition than of logico-scientific cognition. Finally, (d) some of the findings obtained in two recent experimental studies on symbolization will be discussed in order further to support the main thesis of this paper.

From the examples cited by Laffal as illustrations of "opposite speech," there is at least the possibility that these investigators have subsumed under a common rubric essentially dissimilar phenomena with respect to cognitive processes. Thus, when they state that the patient remarked, "I'm hungry," and then refused to take food from his mother, there is at least the possibility that his refusal was based on some delusion with respect to food coming from his mother (cf. 15, p. 40); the refusal does not necessarily warrant the interpretation that the patient really meant "I'm not hungry." Again, when a male schizophrenic, to the statement, *I am a man* responds with "no" (written) and "Nope, I'm no man" (oral), it does not follow that he is confusing or interchanging "no" with "yes"; he may well be deluded with respect to his sex. There are several other examples whose classification involves the same arbitrariness of interpretation, but the above may suffice to ground the criticism that several diversely mediated phenomena may well be grouped by Laffal *et al.* under a single heading.

Leaving aside these questionable instances, there were a number of illustrations provided by Laffal which appear clearly to exemplify "opposite speech" in the sense of involving either (a) a lack of differentiation in the use of affirmative and negative forms of judgment, e.g., "I know," "I don't know," etc., or (b) an undifferentiatedness in linguistic expression for logically opposite meanings, e.g., the use of "like" for both "like" and "hate," etc.

Thus, it seems that "opposite speech" was operative when the patient elaborated on "limping around" by adding "able to walk" or on "disgusted" by remarking "when you're feeling good, happy and contented." Again, it seems that "opposite speech" was in play when the patient completed the sentence *What makes me angry...* with "I feel good, I get very nervous inside, I start to like."

The underlying confusions between affirmation and negation or between opposing predicates seem especially clear in those instances when the patient appeared to be struggling to clarify his thought-expression; for example, when he completed the sentence, *I feel like cursing when...* with "things go good, or if I have any hard luck, get in trouble in any way. Well, if things go right... If I

have some plans and things come up to break it... don't come up to break it."<sup>3</sup>

The grounds on which developmental theorists expect opposite speech phenomena to be more characteristic of primitive thought than of relatively advanced, logico-scientific thought may now be briefly considered. As has been mentioned, thought is defined as more primitive the more undifferentiated and unintegrated it is. Opposite speech is presumed to arise because the content to which the linguistic signs refer is in such a global, undifferentiated state that the linguistic vehicles (seemingly discrete) really share the same global referent (3, pp. 182 ff.; 17, p. 18; 19, p. 40; 20, pp. 119 ff.). Concretely, a linguistic sign like "hate" and a sign like "love" both refer to a sphere of "strong affective response" as distinguished from "neutral reaction"; signs like "hot" and "cold" both refer to a semantic sphere of "extreme deviations from normal temperature," etc. It is due to the undifferentiated character of the referent, then, that two signs or linguistic sign complexes, which in mature thought function to articulate and differentiate a global semantic sphere, in primitive thought are used interchangeably. It follows from this characterization that when there is complete and systematic exchange of symbols ("love" used always for "hate" and vice versa), opposite speech is not necessarily manifest. Opposite speech is in play only where the same sign or sign-complex refers now correctly, now incorrectly.

Turning now to the five domains mentioned above, one may examine the evidence to determine whether there is warrant in concluding that such verbal interchanges or undifferentiated symbolizations are more characteristic of these domains than of logico-scientific domain.

First, even in normal adults conditions of stress or anxiety—presumably leading to a dedifferentiation of thought—promote the emergence of "opposite speech" phenomena, e.g., "left" is said for "right," "on" for "off," etc. Many phenomena subsumed by psychoanalysts under the concept of "psychopathology of everyday life" are clearly cases of opposite speech. In this connection Freud remarks, "Numerous observations have taught

<sup>3</sup> It may be noted that in most of the examples provided by Laffal *et al.*, the manifestations of "opposite speech" may have come about through the omission by the patient of "not" or "don't"; in other words, what often appears to be an interchange of opposites may rather be a positive judgment, indifferently signifying both an affirmative and a negative judgment. There are some instances, however, which seem unambiguously to involve an interchange of opposites, e.g., the patient completes the sentence, "When people make fun of me..." with "I start to like."



me . . . that we frequently interchange contrasting words; they are already associated in our speech consciousness; they lie very close together and are easily incorrectly evoked" (9, p. 43).

Second, opposite speech phenomena<sup>3</sup> are characteristically exhibited in the relation between manifest and latent dream contents. Freud, summarizing his observations of dreams, remarks: "The way in which dreams treat the category of contraries and contradictories is highly remarkable. It is simply disregarded. 'No' seems not to exist so far as dreams are concerned. They show a particular preference for combining contraries in a unity, or for representing them as one and the same thing. Dreams feel themselves at liberty, moreover, to represent any element by its wishful contrary; so that there is no way of deciding at a first glance whether any element that admits of a contrary is present in the dream thoughts as a positive or as a negative" (10, p. 318). If one accepts Freud's inquiries as representative and his analyses as correct, it would seem that opposite speech phenomena are characteristic of the dream state.

Third, it has been argued that there is a strong tendency toward opposite speech in early forms of language. Thus Karl Abel (1, pp. 225 ff.) claimed that in the most ancient languages there characteristically appears the use of a single verbal symbol to represent what for Western adult mentality are two contrasting significations, two contraries at the extreme ends of a series of qualities or activities. Even when distinct terms for the two contraries emerge, their kinship is reflected in the fact that the differentiation is achieved by slightly modifying the "mother form" to express one of the contraries, while the original form is retained for the other. Abel illustrates his thesis with reference to the Coptic language: In Coptic, *kef* is used to express both "to take" or "to let lie"; *ken* designates "strong" and/or "weak." The symbols for "stand" and "move" are *men* and *menmen*, respectively—this last example illustrating the primal emergence and representation of more precise determinations of the common, global sphere.<sup>4</sup>

Research on some contemporary "primitive" (cf. 27) languages yield the presence of similar phenomena. For example, Westermann, in his study of the Ewe language (cited in 6) observes that one and the same adverb serves to designate both "yesterday" and "tomorrow"; again, in the

Shambala language, the same word refers both to the earliest time and the distant future.<sup>5</sup>

Fourth, in their renowned work, *Die Kindersprache*, Clara and William Stern (21) cite many examples of the use of a single symbol among children for what to normal adults are contrasting significations. For example, their daughter, Hilde at 12½ months used "papa" to designate both her father and mother; at 16 months she used "mama" to designate both her mother and father; at 18 months she interchanged "ja" and "nein" (pp. 29); at 21 months she confused "winter" and "sonne," and also "kalt" and "heiss" (pp. 35 ff.); at two years of age, she would use the adverb of place, "zu," for "zu" and "auf" (pp. 56).

According to the Sterns, the use of antithetical words belongs almost exclusively to the first stages of child language. They differentiate a small number of groups of words which seem to be interchanged with some regularity among the children discussed in the literature. One group, for example, consists of place adverbs of motion. Among German children, the words "an" and "auf" are especially interchanged. A second group involves expressions of activity. For instance, one investigator's daughter used "bi" both for being dressed and being undressed. Another investigator observed that his son at 23 months used the word "moga" for "I can" and "I can't," and also "znam" for "I know" and "I don't know." A third group in which antithetical meanings are expressed by a single verbal symbol comprises adjectives, very frequently adjectives of temperature; the Sterns mention that seven of the children discussed in the literature interchanged "heiss" and "kalt," etc.<sup>6</sup>

Decroly's (8, p. 234) investigations supplement the findings cited by the Sterns and others. In studying his child, E., Decroly found many instances of "opposite speech." Thus, *feme* was used to express both "fermer" (close) and "ouvrir" (open); *cho* signified both "chaud" (warm) and "froid" (cold); *lume* expressed both "allumer" (light up) and "eteindre" (extinguish); *pen* was

<sup>5</sup> Cassirer, in discussing this lack of differentiation between expressions for past and future, observes that it derives, in part, from the primitive intuition of time as a *thing*. A sharp distinction is made early between the contents of the now and the not-now, but the not-now is not given any further determination (6, p. 220).

<sup>6</sup> The Chamberlains (7), not cited by the Sterns, give some examples, analogous to those quoted by Laffal *et al.*, with respect to their patient's interchange of "any" for "no" or "none." At 31 months, their girl often used "any," "anything," or "anybody" in the sense of "none," "nothing," "nobody"; thus, she remarked, "Any more pickas in zat book" for "No more pictures . . ." etc.

<sup>4</sup> According to Abel, the same tendency to express contrasting significations by a single symbol characterizes early stages of the Semitic and Indo-European languages. For a more recent linguistic discussion of such undifferentiated, "neutral" symbols, see (13).

used for "prendre" (take) and "donner" (give), etc. All of these findings serve to support the thesis that opposite speech is relatively more characteristic of the linguistic activity of younger children than of older children and adults.

Fifth, students of the language behavior of aphasics have often remarked the occurrence of phenomena subsumable under the concept of "opposite speech." One of the earliest investigators, Bastian, cites several cases in which aphasic patients could say both "yes" and "no," but could not always use these terms appropriately (4, pp. 97, 99). Additional examples of such linguistic interchanges may be found in the works of Head (14), Goldstein (12), and Weisenburg and McBride (23).

Head remarks of one of his patients that "he could not grasp the exact meaning of 'to' and 'past' (the hour), 'high' and 'low,' 'up,' and 'down,' 'back' and 'front,' 'right' and 'left' when contrasted with one another. Asked, 'Would you like to stop in or go out?' he would reply, 'Either,' in order to hide his want of complete comprehension" (14, 1, p. 180).

One of the patients examined by Weisenburg and McBride answered "yes, yes" to all questions when he meant "no" (p. 315). On a test for the employment of opposites, this patient, according to the authors, showed "an increasing tendency to lose the controlling idea 'opposite of.'" Another patient tended to use "only the words 'no' and 'yes' and to say 'no' very decidedly sometimes when she obviously meant 'yes'" (23, p. 557). Goldstein, whose conceptual framework is closely related to that presented here, not only observed such linguistic interchanges in his aphasic patients, but pointed out its analogy to child linguistic behavior (cf. 12, p. 39) as has been attempted here with schizophrenics. Assuming that these examples of "opposite speech" in aphasics are more representative of that group than of normals, this provides another source of evidence for assigning "opposite speech" to a relatively lower stratum in the development of linguistic thought organization.

Two experimental studies on symbolization and thought which are relevant to the problem of "opposite speech" may now be discussed. In one part of the first study, carried out by R. Pollack (18), three groups of children (ranging in age from six years, seven months to 12 years, six months) and a group of college students were presented with the task of selecting colors for the representation of 20 mood terms (e.g., happy, sad; nice, nasty, etc.) so chosen that the total list consisted of 10 pairs of opposites, all subsumable under two broad classes of positively toned and negatively toned words. In general, it was found that younger children tended more than older

ones to employ the same color to represent what, to adults, are more or less antithetical meanings. Moreover, when the opposition between terms was highlighted by juxtaposing the antonyms, the older children and adults seemed to benefit from the experimental polarization of opposites by more consistently selecting different colors for the antithetical words; the younger children showed no noticeable improvement. The groups were then asked to symbolize the mood terms expressively by drawing linear symbols. Only the adults expressively distinguished positive and negative mood terms by linear schematization when the mood terms were presented in random order. When presented with a juxtaposition of opposites, the youngest group of children alone showed no improvement in the tendency to differentiate the positive and negative opposites. This study points to the relative undifferentiatedness among younger children of what later are given determination as opposites.

The second experimental study, carried out by A. Goldman (11), was concerned with symbolic reception and expression among schizophrenics contrasted with normal adults. In the first part of the study, the two groups of subjects were randomly presented with 20 mood terms to be represented in line patterns; again the total group of words consisted of 10 pairs of opposites. In the second part of the experiment, the subjects were presented with line symbols constructed by the experimenter and asked to match the symbols with the mood terms. Goldman found that schizophrenics, both receptively and expressively, were more likely than normals to utilize the same symbol to express antithetical meanings or sharply contrasting symbols to express the same meanings. Thus, a schizophrenic was significantly more likely than a normal to take on one occasion a curved, light, upward line, and on a second occasion an angular, dark, downward line as expressive of "joy" or to use the light, curved, upward line on one occasion as expressive of "joy" and on a second occasion as expressive of "sadness."

The findings of the experiments by Pollack and Goldman, and the general tenor of the evidence from the five domains in which primitive thought expression might plausibly be thought to obtain, suggest that the processes underlying "opposite speech" occur not only in a few schizophrenics employing a strange means to avoid anxiety or to express hostility, but may be found in any individual, characteristically or momentarily operating under conditions conducive to a primitivization of the level of symbolic articulation and organization of experience. It is hoped that this discussion shows the value of a comparative-developmental analysis of the formal aspects of cognitive functioning as a



complement to dynamic considerations for a fuller understanding of schizophrenic symptomatology (cf. 2, Part III).

## REFERENCES

- ABEL, K. *Linguistic essays*. Boston: Houghton Mifflin, 1882.
- ARIETI, S. *Interpretation of schizophrenia*. New York: Brunner, 1955.
- BALDWIN, J. M. *Thought and things*. Vol. 1. New York: Macmillan, 1906.
- BASTIAN, H. C. *Aphasia and other speech defects*. London, 1898.
- BAUDOUIN, C. La regression et les phénomènes de recul en psychologie. *J. de Psychol.* 1928, 25, 795-823.
- CASSIRER, E. *The philosophy of symbolic forms*, Vol. I (trans. by R. Manheim). New Haven: Yale Univer. Press, 1953.
- CHAMBERLAIN, A. F., & CHAMBERLAIN, I. C. Studies of a child. *Ped. Sem.*, 1904, 11, 249-263.
- DECROLY, O. *Comment l'enfant arrive à parler*, Vol. II. Belgium: Cahiers de la Centrale, 1934.
- FREUD, S. *Psychopathology of everyday life*. New York: Mentor, 1952.
- FREUD, S. *The interpretation of dreams* (trans. by J. Strachey et al.). London: Hogarth Press, 1953.
- GOLDMAN, A. E. An investigation of symbolic representation in schizophrenia. Unpublished doctoral dissertation, Clark Univer., 1955.
- GOLDSTEIN, K. *Language and language disturbances*. New York: Grune & Stratton, 1948.
- HATZFELD, H. *Leitfaden der vergleichenden Bedeutungslehre*. Hueber: Munich, 1928.
- HEAD, H. *Aphasia and kindred disorders of speech*. 2 vols. Cambridge: Cambridge Univer. Press, 1926.
- HILL, L. B. *Psychotherapeutic intervention in schizophrenia*. Chicago: Univer. of Chicago Press, 1955.
- LAFFAL, J., LENKOSKI, L., & AMEEN, L. "Opposite Speech" in a schizophrenic patient. *J. abnorm. soc. Psychol.*, 1956, 52, 409-413.
- MESSER, A. Experimentelle psychologische Untersuchungen über das Denken. *Archiv. f. d. ges. Psychol.*, 1906, 8, 1-224.
- POLLACK, R. H. A genetic study of intuitive word meanings. Unpublished doctoral dissertation, Clark Univer., 1953.
- RIBOT, T. *Essay on the creative imagination* (trans. by A. baron). Chicago: Open Court, 1906.
- SIGWART, C. *Logic*. Vol. I (trans. by H. Dendy). New York: Macmillan, 1895.
- STERN, CLARA, & STERN, W. *Die Kindersprache*. Leipzig: Barth, 1907.
- TOULMIN, S. *The philosophy of science*. London: Hutchinson, 1953.
- WEISENBURG, T., & MCBRIDE, K. *Aphasia*. New York: The Commonwealth Fund, 1935.
- WERNER, H. *Comparative psychology of mental development*. (3rd ed.) New York: International Universities Press, 1957.
- WERNER, H. Microgenesis and aphasia. *J. abnorm. soc. Psychol.*, 1956, 52, 347-353.
- WERNER, H. The concept of development from a comparative and organismic point of view. In D. Harris (Ed.), *The concept of development*. Minneapolis: Univer. of Minnesota Press, 1957.
- WERNER, H., & KAPLAN, B. The developmental approach to cognition: its relevance to the psychological interpretation of anthropological and ethnolinguistic data. *Amer. Anthropol.*, 1956, 58, 866-880.
- WHYTE, L. L. *Accent on form*. New York: Harper, 1954.

Received January 28, 1957.

## ETHNOCENTRISM AND STIMULUS GENERALIZATION: A REPLICATION AND FURTHER STUDY

FRANKLYN N. ARNHOF

*Mental Health Research Unit, Dept. of Mental Hygiene, Syracuse, N. Y.*

IN A previous paper (2), the writer reported an exploratory study in which ethnocentrism test scores and error scores on a visual-spatial task of stimulus generalization (SG) were found to be positively related ( $r = .44$ ,  $N = 25$ ) tending to give support to the hypothesis that ethnocentrism represents an "overgeneralization phenomenon." During the course of an extensive investigation of SG, the previous study was replicated with refinements and additional controls. In addition to the visual-spatial task (lights), a more cognitive task (blocks) was added to the

present study and both SG tasks were related to the subjects' (Ss') scores on the California Ethnocentrism Scale (E scale) (1). As this study was part of a larger study reported elsewhere (3), and the two generalization tasks have been previously described in detail (2, 3, 4, 5), the reader is referred to other sources for descriptions of the tasks and the experimental design.

The volunteer Ss were 60 female student nurses, aged 18 through 22. As far as can be determined, the sample was comparable to the one in the previous study (2).

## RESULTS AND DISCUSSION

Following administration of the two SG tasks, Ss were given the E scale and the CVS Intelligence Scale. Prior to the study proper, all Ss had been administered the group form of the Minnesota Multiphasic Personality Inventory (MMPI).

Generalization was defined as the number of errors made on the two tasks. Total error scores were computed for each S on each task. These were then correlated with Ss' E scale scores, yielding  $r$ 's of .10 (lights) and .15 (blocks), values falling far short of significance.

IQ was not found to be a significant factor in SG error scores, although its relation to the E scale ( $r = -.31$ ) was consistent with previously reported findings (1).

Test-retest scores on the E scale were found reliable ( $r = .85$ ) and overdefensive responding as measured by the MMPI K scale was only minimally related to E scale scores ( $r = -.29$ ), ruling out these factors as explanation for the present failure.

The previous positive findings must therefore probably be attributed to chance. While present results do not preclude the possibility that ethno-

centrism is related to overgeneralization in thinking, the functions involved in performance on the tasks investigated do not appear to be on the same psychological continuum.

## REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ARNHOF, F. N. Ethnocentrism and stimulus generalization. *J. abnorm. soc. Psychol.*, 1956, **53**, 138-139.
3. ARNHOF, F. N., & LOY, D. L. The relationship between two measures of stimulus generalization: The influence of intelligence upon performance. *Psychol. Rep.*, in press.
4. BROWN, J. S., BILODEAU, E. A., & BARON, M. P. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52-61.
5. BUSS, A. H. Stimulus generalization as a function of clinical anxiety and direction of generalization. *J. abnorm. soc. Psychol.*, 1955, **50**, 271-273.

Received November 8, 1956.

## SOCIAL DESIRABILITY AND PROBABILITY OF ENDORSEMENT OF ITEMS IN THE INTERPERSONAL CHECK LIST<sup>1</sup>

ALLEN L. EDWARDS

*University of Washington*

ONE OF the most widely used methods for obtaining self-descriptions of personality is the personality inventory. As is well known, the typical personality inventory consists of a number of statements or items to which a person is asked to respond by indicating all those that he believes characterize himself. His responses to the statements thus provide his self-description. If a large number of individuals are given the same inventory, it is possible to find the proportion endorsing each statement, that is, the proportion who say that a given statement is applicable in self-description. Each of these proportions may be regarded as the probability that a given statement will be answered "Yes," "Agree," or "True," depending on the form of the inventory.

It has been found that it is possible to predict these probabilities of endorsement with a rather

high degree of accuracy. The predictions are possible because the probabilities of endorsement are linearly related to another independently determined characteristic of the statements, namely, their social desirability scale values. The relationship between these two variables has been found consistently for a variety of statements investigated.

The social desirability scale value of a personality statement refers to its location on a psychological continuum of social desirability. A judging group, for example, is asked to make comparative judgments of the degree of social desirability of each of the statements in a set. On the basis of these judgments, and by means of one of the psychological scaling methods, a scale value is obtained for each statement on a psychological continuum of social desirability. High scale values on this continuum represent statements that are judged as socially desirable and low scale values represent statements that are judged as socially undesirable.

Suppose we scale the statements in a given

<sup>1</sup>This paper was presented at the American Psychological Association meetings, Chicago, Illinois, September 1, 1956. The study was supported by a grant from the Graduate School of the University of Washington.



personality inventory for social desirability in terms of the judgments of one group of subjects. The inventory is then administered to a new and independent group, that is, one other than that used in determining the social desirability scale values, with the usual instructions to provide self-descriptions. For this second group, the proportion endorsing each of the statements in the inventory is obtained. It is then possible to determine the relationship between probability of endorsement and social desirability scale value for the set of statements.

Following the procedure just described, Edwards (1) found a product-moment correlation of .87 between probability of endorsement and social desirability scale value for a set of 140 personality statements. This finding has since been substantiated by Kenny (4) and by Hanley (3) and, indirectly, by Rosen (10). Edwards' research was done with statements subsequently used in developing the PPS (2), a forced-choice inventory in which statements are paired on the basis of their social desirability scale values. Kenny's research was done with a set of personality statements originally used in a study by Zimmer (12). The statements used by Hanley were selected from three of the MMPI scales. Rosen's study was also done with statements from the MMPI.

In all of the studies mentioned above, the subjects given the personality inventories were asked to sign their names on their answer sheets. That this requirement may be an important factor in the relationship between probability of endorsement and social desirability scale value is indicated by Meehl and Hathaway (8). After surveying research by Maller (7), Olson (9), and Spencer (11) on self-descriptions made under anonymous conditions, Meehl and Hathaway conclude that requiring subjects to sign their names to their self-descriptions has been shown to have a definite influence on the descriptions obtained. The present study was undertaken, therefore, to determine whether the relationship between probability of endorsement and social desirability scale value would be found when subjects made their assessments anonymously.

#### METHOD

The statements used in the study were those appearing in the 128-item Interpersonal Check List (ICL). The ICL was developed by staff members of the Kaiser Foundation Research Project in Psychology. It has been described in an article by LaForge and Suczek (5) and has been prepared in booklet form by Leary (6). The ICL can be used for purposes of both self-assessment and assessment by others. In self-assessments, the S checks each item he regards as self-descriptive. The ICL thus closely resembles the typical personality in-

ventory in which the S responds "Yes" or "No" to the individual items in the inventory. The ICL statements were selected for investigation because it seemed of value to extend our knowledge of the relationship between probability of endorsement and social desirability scale value with respect to new and different statements which had not been used in any of the previous studies.

The 128 items in the ICL were given to a group of 58 male and 39 female college students who were asked to make comparative judgments of the social desirability of each of the items. Judgments were made in terms of a nine-interval scale, the first interval being defined as highly socially undesirable and the last as highly socially desirable, following the same instructions used in an earlier study (1). Using these judgments, a scale value for each item can be obtained by the method of equal-appearing intervals.

Before actually calculating scale values of the items, the distributions of judgments for the males and females were examined to determine whether the judgments obtained from the two groups could be pooled. This was done by determining the interval in which the scale value of each statement would fall for each group of judges considered separately. The correlation between these scale intervals over the 128 statements was .90, a value sufficiently high to warrant pooling the two distributions. On the basis of the judgments of the combined group of 94 to 97 judges, equal-appearing interval scale values were found for the 128 items. The scale value for a particular item corresponds to the median of the distribution of judgments for the item on the nine-interval scale. The number of judges varies for the items because of failure to obtain complete sets of judgments for all 58 male judges.

The ICL was then given to students registered in speech classes at the University of Washington with standard instructions for obtaining self-descriptions, except that they were told *not* to put their names on their test booklets. In addition, the students were told that their instructor was not interested in identifying anyone's answers and that the data were to be used in a study of sex differences. It was hoped that these instructions would offer some assurance to the students that their answers would indeed remain anonymous.

#### RESULTS

Complete records were obtained from 67 female and 66 male students. For each of these two groups the frequency of endorsement of each item was found and transformed to a proportion. These proportions were regarded as the probabilities of endorsement for the various items. The product-moment correlation between the female and male probabilities over the set of 128 items was .95. Because of the high relationship between the two sets of probabilities, the data were pooled and a new set of probabilities obtained for the combined group of 133 Ss.

Probability of endorsement of an item was then plotted against the previously determined social desirability scale value of the item. The plot was linear and the product-moment correlation between the two variables was .83.

## DISCUSSION AND CONCLUSIONS

If anonymity has any influence on self-descriptions, it would seem that the influence should be such as to produce more frank evaluations. Under anonymous conditions, therefore, one might expect responses to personality statements to be less related to the social desirability scale values of the statements than when Ss are asked to sign their names to their self-assessments. The correlation of .83 obtained between the two variables of interest in the present study is, however, not significantly different from the value of .87 reported previously for a different set of statements and a different sample of Ss and where the assessments were not made anonymously.

The correlation reported, of course, does not bear upon the problem of whether or not we would have obtained a significantly higher correlation for the same set of statements if Ss had been asked to sign their names to their self-descriptions. More important would seem to be the fact that the correlation obtained in the present study is of sufficient magnitude to indicate that assurance of anonymity does not eliminate nor drastically change the nature of the relationship previously found between probability of endorsement and social desirability scale value.

## REFERENCES

1. EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability

- that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90-93.
2. EDWARDS, A. L. *Manual for the Personal Preference Schedule*. New York: Psychological Corp., 1954.
3. HANLEY, C. Social desirability and responses to items from three MMPI scales: D, Sc, and K. *J. appl. Psychol.* 1956, **40**, 324-328.
4. KENNY, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. *J. consult. Psychol.*, 1956, **20**, 315-318.
5. LAForge, R., & SUCZEK, R. F. The interpersonal dimension of personality: III. An interpersonal check list. *J. Pers.*, 1955, **24**, 94-112.
6. LEARY, T. The Interpersonal Check List. Berkeley, Calif. Psychol. Consult. Service. (Undated)
7. MALLER, J. B. The effect of signing one's name. *Sch. and Soc.*, 1930, **31**, 882-884.
8. MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525-564.
9. OLSON, W. C. The waiver of signature in personal reports. *J. appl. Psychol.*, 1936, **20**, 442-450.
10. ROSEN, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, **52**, 151-158.
11. SPENCER, D. Frankness of subjects on personality measures. *J. educ. Psychol.*, 1938, **29**, 26-35.
12. ZIMMER, H. Self-acceptance and its relation to conflict. *J. consult. Psychol.*, 1954, **18**, 447-449.

Received January 14, 1957.

SOCIAL SETTING AND CONFORMITY TO A LEGAL REQUIREMENT<sup>1,2</sup>

ABRAM M. BARCH, DON TRUMBO, AND JOHN NANGLE

*Michigan State University*

THE PURPOSE of the study was to note the effect on a subject's behavior of observing others conform or fail to conform to a legal requirement. The specific hypothesis under test was that the signalling behavior of the driver of a motor vehicle preparing to turn at an intersection would be related positively to the signalling behavior of the driver of an immediately preceding motor vehicle.

The research was stimulated by the work of Blake and his associates concerned with the influence of social background or social setting on conformity behavior (1, 2, 3, 4), but the data are also relevant to other formulations of imitation

and social conformity. In the situation studied, there is a minimum of interaction between the subject and the model, and the subjects can be assumed to have had considerable opportunity to learn what the typical behavior of other individuals in situations similar to the test situation would be. Unlike the settings investigated by Blake *et al.*, conformity or nonconformity in the present situation is to a legal requirement.

## METHOD

*The Observations*

Preliminary observations indicated that the variation in turn-signalling present in the naturalistic situation was sufficient for the purpose of the study.

Four different intersections in the Greater Lansing area were observed for a total of 61 hours

<sup>1</sup> The study was supported by the Highway Traffic Safety Center of Michigan State University.

<sup>2</sup> The authors are indebted to Dr. Charles Hanley for statistical advice and assistance.



over a four-week period in the summer of 1956. The particular hours during which the intersections were observed depended on considerations involved in a broader study of turn-signalling behavior; all data were collected during daylight hours and in good weather (no fog or rain). Intersections studied included two instances of two two-lane roads meeting at a two-way stop, one four-lane highway intersecting a two-lane road at a stop signal, and one six-lane divided highway intersecting a three-lane road at a stop signal.

During any period of observation, the observer noted all cars moving from a given direction on a given leg of the intersection and recorded the following information for all turning cars: (a) presence or absence of a signal to turn; (b) direction of turn; (c) presence of another car 100 feet or less behind the turning car when it began its turn; (d) sex of the driver. (At two of the seven legs observed, only cars turning in a given direction were recorded because of the low occurrence of turns in one direction and/or the difficulty of accurately noting turns in both directions.)

The observer stationed himself near the intersection where he could best make the observations required. In most cases he was diagonally across the intersection from the observed traffic movement. At other times he stood about 100 feet ahead of the intersection along the lane(s) being observed. Data collected under control conditions designed to eliminate the possibility that drivers might note the presence of observers indicated that the visible presence of observers did not affect the percentage of cars signalling for a turn. In a reliability check, agreement between two observers for the same cars was 99 per cent or better for all categories except "the presence of a following car" (94 per cent).

During the periods of observation, 4,229 male drivers and 1,004 female drivers of passenger cars turned at the locations under study.

#### *Criteria for Scoring Conformity*

Michigan state law requires that intention to turn be signalled by either hand signal or electric signal but does not specify the distance this signal must be given prior to turning or make a distinction between a signal for a left turn and one for a right turn. Therefore, a driver was designated as signalling if he blinked his left or right turn signal light or gave any hand signal, except a hand signal for stopping, regardless of the direction of turn.

#### RESULTS

The first step in the analysis consisted of extracting the appropriate pairs from the field observation data. The criteria for selection were: (a) both cars turned at the intersection under

TABLE 1  
REACTIONS OF SUBJECTS TO SIGNALLING (S)  
AND NONSIGNALLING (NS) MODELS

Behavior of Model		Behavior of Subject							
		High Conformity Legs				Low Conformity Legs			
		Right Turn		Left Turn		Right Turn		Left Turn	
		S	NS	S	NS	S	NS	S	NS
S	n	140	74	167	62	48	100	77	55
	%	65	35	73	27	32	68	58	42
NS	n	84	64	36	21	42	103	64	58
	%	57	43	63	37	29	71	52	48

study; (b) the following car was 100 feet or less behind the lead car when the lead car began its turn; and (c) no cars were in the same lane between the two turning cars. A given car could appear in only one pair of cars; i.e., it could not be a following car for one pair and the lead car for the next pair. The total number of such pairs was 1,195. Of these, 723 pairs had a conforming model and 472 pairs had a nonconforming model.

The signalling behavior of the follower was significantly related at the .01 level to the signalling behavior of the lead car ( $\chi^2 = 15.78, 1 \text{ df}$ ). This positive relationship, although highly significant, was found to be a weak one. The strength of the association as estimated by tetrachoric  $r$  was .19.

A future analysis of the data was felt desirable since the general level of conformity (percentage of all turning cars signalling without regard to presence of a model car) varied with the direction of turn and the leg of the intersection. Women also tended to signal more frequently than men. However, the difference was small and the pattern of relationships within the data for women was quite similar to that for men. Therefore, men and women drivers were combined in all comparisons.

Table 1 presents the number and percentage of subjects signalling or failing to signal in relation to the signalling behavior of the model under various conditions of direction of turn of the subject and the general conformity level characteristic of the intersection leg at which the turn was observed. General conformity level for an intersection leg was taken as the percentage of turn-signalling for all cars turning in a given direction without regard to the presence of a model car. Intersection legs were classified as high conformity if the percentage of turning-signalling for left turns was between 62.5 and 88 and the percentage of signalling for right turns was between 37.5 and 62.5. Low conformity legs had a signalling percentage between 37.5 and 62.5 for left turns and one between 20 and 37.5 for right turns.

Table 2 presents the results of the  $\chi^2$  analyses

TABLE 2  
STATISTICAL ANALYSES OF THE INFLUENCE  
OF THE MODEL

Respondent Categories	$\chi^2$ <sup>a</sup>	<i>p</i>	<i>r</i> <sub>tet</sub>
All cases	15.78	.01	.19
All left turns	6.62	.05	.20
All right turns	4.82	.05	.14
All cases at high conformity legs	6.75	.05	.18
All cases at low conformity legs	1.17	—	.09
Left turns at high conformity legs	1.67	—	.18
Right turns at high conformity legs	2.43	—	.14
Left turns at low conformity legs	0.66	—	.09
Right turns at low conformity legs	0.27	—	.07

<sup>a</sup> All analyses have 1 *df*.

of the influence of the signalling behavior of the model on the subject's signalling for various combinations of conditions. Tetrachoric *r* is provided for each combination to show the strength of the association. In all comparisons the relationship was positive to the hypothesis of influence. The strength of the relationship was low for all analyses, ranging from .07 to .20.

An attempt was made to increase the possible identification of the subject with the model by considering only those pairs in which both members of the pair were of the same sex and turned in the same direction. There were 691 such pairs. The  $\chi^2$  for this analysis, as would be expected, was significant at the .01 level ( $\chi^2 = 7.33$ , 1 *df*). However, the tetrachoric *r* was found to be .17,

essentially the same as that obtained for the other comparisons.

The results supported the hypothesis that the behavior exhibited by others in a situation involving a legal norm influences a subject's behavior. The weakness of the demonstrated influence should be considered in light of the lack of formal relationships or of interaction, except in a minimal sense, between the model and the subject and the subject's ample opportunity to observe the behavior of others in the same or similar situation prior to the test observation.

#### SUMMARY

Observations made under normal traffic conditions were analyzed to determine the influence on the signalling behavior of automobile drivers of the signalling behavior of other drivers. A weak but significant positive relationship was found between observation of conformity of others to a legal requirement and self conformity.

#### REFERENCES

1. BLAKE, R. R., & MOUTON, JANE S. Present and future implications of social psychology for law and lawyers. *Symposium Issue, Emory Univer. J. Public Law*, 1955, 3, 352-369.
2. FREED, A. M., CHANDLER, P. J., MOUTON, JANE S., & BLAKE, R. R. Stimulus and background factors in sign violation. *J. Pers.*, 1955, 23, 499.
3. LEFKOWITZ, M., BLAKE, R. R., & MOUTON, JANE S. Status factors in pedestrian violations of traffic signals. *J. abnorm. soc. Psychol.*, 1955, 51, 704-706.
4. ROSENBAUM, M., & BLAKE, R. R. Volunteering as a function of field structure. *J. abnorm. soc. Psychol.*, 1955, 50, 193-196.

Received January 16, 1957.

## PUBLIC AND PRIVATE CONFORMITY UNDER DIFFERENT CONDITIONS OF ACCEPTANCE IN THE GROUP<sup>1</sup>

HERBERT MENZEL

Columbia University

DITTES and Kelley (2) recently reported an experiment on conformity to group norms which showed, among other things, that the subject's experimentally induced feeling of acceptance within the group had a rather different effect on his private and public expressions of conformity

<sup>1</sup> This paper may be identified as Publication No. A-236 of the Bureau of Applied Social Research, Columbia University, and is an outgrowth of a study conducted by James Coleman, Elihu Katz, and the present writer.

to the group's previously established judgment. It was found, in brief, that individuals whose feeling of being accepted in the group was high felt the greatest freedom to express any disagreement with the group's judgment publicly. Individuals whose feeling of acceptance was low exhibited high public conformity but showed little conformity in their private expressions. Individuals who were made to feel an "average" level of acceptance exhibited a "high degree of genuine adherence to the norms . . . extending even to conditions of privacy." The



present note addresses itself exclusively to the similarity or difference between public and private conformity, and not to the determination of private conformity as such.

#### DATA

The conclusions of Dittes and Kelley in this respect can be applied in slightly modified form to data, originally secured for another purpose (see 1, 4), on the use of new drugs by physicians. Eighty-five per cent of the general practitioners, internists, and pediatricians practicing in four selected cities were interviewed regarding their use of a certain widely used family of drugs, comprising several new variants which had made their appearance successively in the course of recent years. At the same time, the records of prescriptions written by these same doctors were examined in the local pharmacies for sampling dates extending over a period of 15 months. One may think of the physician's writing of prescriptions as his private behavior, and of his declarations in the course of an interview as the analogue of "public" expressions in the sense of the Dittes and Kelley experiment. So defined, "public" and "private" data are available on three items of behavior: (a) the date on which the doctor used "gammanym," the newest variant of the family of drugs, for the first time; (b) which of the three chief available variants he favored at the time the interviews were conducted; (c) which of the two chief variants then available he favored 15 months earlier.

#### RESULTS AND DISCUSSION

##### *Existence of Norms*

In the fluid situation created by the recurrent appearance of new drugs, there is no established norm on record as to which of two or three competing variants a physician should prefer, nor as to the "right" date on which to introduce a new drug into one's practice. What the norm in these respects was at the time the interviews were conducted can, however, be inferred from the prevailing direction of the discrepancies between private performance and public accounts of behavior: (a) One-half of the physicians stated as the date on which they first used the newest drug a date earlier—often many months earlier—than that established by the prescription search (Table 1).<sup>2</sup> (b) Two-thirds

<sup>2</sup> The extent of this discrepancy can hardly be accounted for by the sampling of days in the prescription search. (Three consecutive days per month were sampled.) It is true that the presumed starting date of gammanym use, obtained by recording prescriptions on only a sample of days, can err only by being too late, and not by being too early. It is, however, possible to estimate the size of that error, since not only prescriptions for the new drug were recorded, but also prescriptions for its predecessors. It was found that during the

TABLE 1  
DATE OF INTRODUCTION OF A NEW DRUG:  
INTERVAL BETWEEN INTERVIEW STATEMENTS  
AND PRESCRIPTION RECORD

Date According to Interview	Percentage of Doctors (N = 70)
Follows prescription date by 2 months or more	24%
Falls within one month's distance of prescription date	27
Precedes prescription date by 2-7 months	25
Precedes prescription date by more than 7 months	24
	100%

TABLE 2  
DRUG VARIANT FAVORED AT TIME OF INTERVIEW

Variant Favored According to Interview Statement	Variant Favored According to Prescription Record		
	Oldest	Middle	Newest
Oldest	33%	8%	6%
Middle	25	32	18
Newest	42	60	76
Total N	100% (24)	100% (25)	100% (61)

of the physicians whose prescription record at the time of interviewing showed predominant use of the oldest drug stated that they preferred one of

average two-month period the proportion of doctors who were missed by the prescription search—i.e., for whom prescriptions for neither the new nor the older drugs were recorded on the sampling days falling within that period—was 23.2% (with a standard deviation of .06%). This may be taken as an estimate of the proportion of doctors who were missed by the prescription search during the two months starting with their respective introduction of the new drug. But since few doctors, if any, made a complete switch from older drugs to newer drugs, one must add to the estimated 23.2% who were altogether missed by the prescription search during these two months the doctors who were covered by the prescription search, but happened to prescribe only older drugs on the sampling days in question. This is estimated at 4.6% of all the doctors in the sample. Thus, if all the doctors had stated their introduction date accurately during the interview, the proportion who would, due to the sampling of days, nevertheless appear to have up-dated themselves by two months or more may be estimated at 23.2% + 4.6% or 27.8%. Actually, as Table 1 shows, 49% appear to have up-dated themselves to this extent.

The above estimate of 4.6% is based on the following considerations. Of all the prescriptions written for this family of drugs by doctors during the sample days within two months after their respective first sampled prescription of gammanym, 67% were for older drugs. The proportion of their prescriptions which ordered the older drugs during the two months following their first

TABLE 3  
DISCREPANCIES IN DATE OF FIRST PRESCRIPTION  
OF A NEW DRUG, BY SOCIOMETRIC STATUS

Introduction Date According to Interview Statement	Number of Designations Received as Advisor		
	None	One or Two	Three or More
Earlier than by prescription record	63%	33%	36%
Same as by prescription record <sup>a</sup>	26	33	21
Later than by prescription record	11	33	43
Total N	100% (35)	100% (21)	100% (14)

<sup>a</sup> Allowing a one-month tolerance.

the newer drugs. Sixty per cent of those whose current prescription record showed predominant use of the middle one of the three drugs stated that they preferred the newest (Table 2). (c) A substantial portion (29%) of the 52 doctors who, according to their prescription record, had prescribed predominantly the oldest variant 15 months before the interviews took place, failed to make any mention of this drug when recounting the history of their use of drugs of this general kind.

From these results it may be concluded that there prevailed, at the time of the interviewing, norms in these communities of physicians which placed high value on (a) having introduced the newest drug early; (b) currently favoring the newest drug over its predecessors; (c) having favored the middle drug over the oldest 15 months earlier (i.e., before the appearance of the newest drug). Thus, with respect to all three items of behavior, the preponderant direction of the discrepancies between the prescription record and the interview statements is such that the interview statements make the doctor appear more "up-to-date." This finding supports the inference that the prevailing norm in these medical communities at the time of interviewing was generally favorable to "modernity," at least in the use of drugs. The

actual prescription of gammanym was probably not far different. The average doctor wrote 7.8 prescriptions for this family of drugs on the sampling days falling within two months of his first sampled prescription of gammanym, and probably as many on the sampling days falling within two months of his first actual prescription of gammanym. The probability of all 7.8 of these prescriptions having been for the older drugs is therefore estimated as the seventh power of .67, which is .06. Applying this factor to the estimated proportion of doctors not altogether missed by the prescription search during the two months starting with their introduction of gammanym ( $100\% - 23.2\% = 76.8\%$ ) yields a figure of 4.6%.

TABLE 4  
DISCREPANCIES IN DATE OF FIRST PRESCRIPTION  
OF A NEW DRUG, IN RELATION TO SOCIOMETRIC  
STATUS, WITH ACTUAL PERFORMANCE  
HELD CONSTANT

Percentage of Doctors who Stated a Drug Introduction Date Earlier (Later) Than the Median Date Stated by Those who Actually Intro- duced the Drug Simulta- neously with Them	Number of Designations received as Advisor		
	None	One or Two	Three or More
Earlier than median	60%	45%	36%
Later than median	40	55	64
Total N	100% (35)	100% (21)	100% (14)

findings to be reported below, however, are of interest even if the above three sets of discrepancies merely express three separate norms relating specifically to the use of the particular drugs in question.

#### *Conformity and Acceptance in the Group*

When the physician's standing in his medical community is taken into account, it is found that the discrepancies in the prevailing direction occur most frequently among the low-status physicians, who tend to make themselves appear more up-to-date than they really are. An individual's standing in the community of his colleagues is measured by the number of designations accorded him in response to the sociometric question: "When you need information or advice about questions of therapy, . . . on whom are you most likely to call?" It is chiefly those seldom or never named in answer to this question whose interview statements make them appear more up-to-date than their prescription record.

Table 3 shows this relationship for the date on which the doctor used the newest of the drugs for the first time. The contrast is quite marked, but is partly due to a ceiling effect: high-status individuals did in fact tend to be more up-to-date in the timing of their first use of this new drug, and hence had less opportunity to deviate, in their interview statements, in the up-to-date direction. However, the phenomenon persists even when actual performance is held constant, as shown in Table 4. Doctors were first divided into two-month intervals according to their drug introduction date as established by the prescription search. Within each such two-month interval, the median interview statement as to the introduction date of the drug was then determined. Each doctor was then classified according to whether his own interview statement gave a date later or earlier than the median statement of doctors who actually introduced the drug within the same two-month inter-



TABLE 5  
DISCREPANCIES IN DRUG FAVORED AT THE  
TIME OF INTERVIEWING, IN RELATION  
TO SOCIOMETRIC STATUS

	Number of Designations Received as Advisor:		
	None	One or Two	Three or More
Percentage who stated they fa- vored a <i>newer</i> variant than prescription rec- ord shows, among those who prescribed predominantly:			
the oldest variant	67% (12)	77% (9)	33% (3)
the middle variant	64% (11)	63% (8)	50% (6)
Percentage who stated they fa- vored an <i>older</i> variant than prescription record shows, among those who prescribed predominantly:			
the middle variant	9% (11)	0% (8)	17% (6)
the newest variant	21% (24)	29% (14)	23% (13)

TABLE 6  
DISCREPANCIES IN DRUG FAVORED 15 MONTHS  
PRIOR TO TIME OF INTERVIEWING, IN  
RELATION TO SOCIOMETRIC STATUS

	Number of Designations Received as Advisor		
	None	One or Two	Three or More
Percentage who stated they fa- vored the newer variant, among those who pre- scribed predom- inantly the oldest	33% (24)	30% (20)	13% (8)

val as he did. It is seen that even when possible ceiling effects are eliminated by holding the date of introduction according to the prescription record constant, those of low standing in the group show a quite disproportionate tendency to report an introduction date which precedes that established from the prescription record.<sup>3</sup>

<sup>3</sup> None of the main results (Tables 4, 5, and 6) are statistically significant in the usual sense, but generalizability of the descriptive results beyond the four cities

Table 5 records the results obtained with respect to the second item of behavior: variant favored at the time the interviews were conducted. The necessity to control for actual performance here makes for very small base figures, and the results are somewhat less consistent than in Table 4, but they are substantially of the same order. Those of high standing in the group consistently "up-date" their behavior less than do those of low standing, although those of intermediate standing do not always fall into the expected place. The two last rows of Table 5 show that this difference is specific to distortions in the direction of the norm; the relatively few distortions that occur in the opposite direction are, if anything, more prevalent among those of high standing.

Table 6 records the results for the third item of behavior: variant favored, among the two then available, 15 months prior to the time of interviewing. Once again, those of high standing in the community of colleagues only seldom reported having used, 15 months before the interviews took place, a newer drug than predominated among their recorded prescriptions, while those of low standing frequently made such reports.

#### DISCUSSION

Thus, with respect to all three items of behavior on which information is available, doctors of low standing in the community frequently reported, during an interview, more up-to-date behavior

surveyed could not be claimed in any case. (For Table 4, collapsed to 2 x 2 format, chi square [corrected for continuity] = 1.721;  $df = 1$ ,  $p = .096$  for a one-tailed test. [See 5, Table 7, p. 125; 6, pp. 108-110].) If the results were based on a sample randomly drawn from a defined and much larger population, observations on the whole population might very well fail to confirm them. In fact, however, the observations were made on 85% of the target population (general practitioners, pediatricians, and internists practicing in four cities). There is therefore little doubt that the findings would be confirmed by observations on the entire target population. On the other hand, neither the sample nor the target population can be claimed to represent any larger population, regardless of the outcome of significance tests.

It is quite another matter that the results are, of course, intended as contributory evidence of the more general applicability of Dittes' and Kelley's conclusions. Establishment of such a proposition ultimately depends on confirmation in a variety of concrete situations (with diverse populations, norms, indicators of acceptance, and expressions of conformity), rather than on the generalizability of one particular finding to a larger population. They are here shown to apply to three items of behavior in one particular situation rather different from the original experiment. The position taken here is essentially that expressed at some length in Lipset *et al.* (3, pp. 427-432).

than is indicated by the prescription record, while doctors of high standing seldom did so. Since, as was shown above, the general norm favored being "up-to-date" in the three matters under investigation (and perhaps in other matters as well), these results seem to be close parallels to the corresponding findings of the Dittes and Kelley experiment.

One may note that the high-status doctors not infrequently deviated in their "public" (i.e., interview) statements from their prescription record in the "wrong" direction, that is, so as to appear less up-to-date than they actually were. It is not clear from Dittes' and Kelley's report whether individuals who agreed with the group privately ever expressed disagreement in public. If not, the contrast may be due to the fact that in the present study the norm of the group was not "crystallized and registered," while the doctors' statements were subject to a certain amount of unmotivated error due to faulty recall, which played no part in Dittes' and Kelley's experiment. It is also possible that there existed, uniquely among the high-status individuals, a competing norm in favor of conservatism; but this seems an unlikely explanation in view of the simultaneous finding that high-status individuals tended to be more up-to-date in fact than low-status individuals, at least in the date of their introduction of the newest of the drugs. (Why high-status individuals were more up-to-date in their actual behavior will be discussed elsewhere. See also [1].)

#### SUMMARY

An experiment by Dittes and Kelley (2) had shown that individuals who felt accepted in their

group felt relatively free to express any disagreement with the group's judgment publicly, while those whose feeling of acceptance was low exhibited much higher conformity in their public than in their private expressions. Data from a survey among physicians provide an analogue in the form of interview statements about prescriptions of new drugs (public expressions) and the doctor's actual prescription record (private behavior). A norm of being up-to-date with respect to three choice situations is inferred from the prevailing direction of discrepancies between private performance and public accounts. Acceptance in the group was measured sociometrically. In all three choice situations, those of low acceptance reported more up-to-date behavior than indicated by the prescription record, while doctors of high acceptance seldom did so.

#### REFERENCES

1. COLEMAN, J. S., KATZ, E., & MENZEL, H. The diffusion of an innovation among physicians. *Sociometry*, in press.
2. DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100-107.
3. LIPSET, S. M., TROW, M. A., & COLEMAN, J. S. *Union democracy*. Glencoe, Ill.: Free Press, 1956.
4. MENZEL, H., & KATZ, E. Social relations and innovation in the medical profession: the epidemiology of a new drug. *Publ. Opin. Quart.*, 1956, **19**, 337-52.
5. PEARSON, E. S., & HARTLEY, H. O. *Biometrika tables for statisticians*. Cambridge: Univer. Press, 1954.
6. SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Received February 1, 1957.

## RESOLUTION OF APPROACH-APPROACH AND AVOIDANCE-AVOIDANCE CONFLICTS<sup>1</sup>

ABE ARKOFF

*University of Hawaii*

CONSISTENT with the deductions of Lewin (5), Hovland and Sears (4) found that avoidance-avoidance conflicts produced more blocking and compromise responses in the manual motor situations they studied than did approach-approach ones. Barker (1), however, found no consistent evidence that choice time and "vicarious trial and error behavior" were greater in children choosing between paired repellent liquids than between

attractive ones. In a second experiment (2), he presented college students with paired hypothetical alternatives involving personal characteristics and environmental conditions and found that the frequency of choices marked "uncertain" was greater for negative than positive alternatives. The present investigation attempted to devise comparable approach-approach and avoidance-avoidance situations that would involve the subjects emotionally and permit clear-cut measures of conflict behavior.

<sup>1</sup> The author wishes to acknowledge the generous assistance of Dr. John M. Digman.



## METHOD

Forty male and 40 female students from the introductory psychology course served as subjects (Ss). Conflicts were created by systematically pairing with one another seven personal characteristics: adjustment, attractiveness, health, intelligence, popularity, talent, and wealth. Two types of pairings were employed, one in which the S was required to designate the alternative that he would rather have in greater degree than he had at present (approach-approach conflict), and a second in which he was required to designate the alternative that he would rather have in lesser degree than he had at present (avoidance-avoidance conflict). In the 42 conflicts presented, 21 described approach-approach situations and 21 described avoidance-avoidance ones. Each choice was typed on a separate 3 x 5 card with the question "Which would you rather be?" centered above, and the alternatives boxed below on left and right portions of the card. An example of an approach-approach conflict follows (for the avoidance-avoidance analogue "less" was substituted for "more"):

Which would you rather be?

More attractive than you are now.	More intelligent than you are now.
--------------------------------------	---------------------------------------

In the first part of the experiment, the experimenter (E) arranged the 42 conflict cards in random order and placed them face down in a single stack before the S who was given written instructions as follow:

Each card describes a conflict which you must resolve. When the experimenter gives the signal, turn up the topmost card and study the conflict presented. If your choice is the alternative on the left, place the card, face down, on the table to the left of the pile. If your choice is the alternative on the right, place the card, face down, on the table to the right of the pile.

Pay no attention to the timing. Take as much or as little time as you like with each card.

Imagine that each conflict *really* confronts you. Be sure that your choice is the one that you would *really* make if you *really* had to decide.

As soon as the S turned up the conflict card (and before he actually began to read the conflict) the E began the timing. Timing was stopped as soon as the card was face down and flat on the table.

The conflict cards were then placed once more in random order, and the S, without timing, was given written instructions as follow:

Divide the cards into two piles of 21 cards each. In one pile place the cards describing the conflicts which you found relatively easier to resolve. In the other pile place the cards describing the conflicts which you found relatively harder to resolve. You may proceed to do this in any fashion you wish.

## RESULTS

Two measures of conflict behavior were available: the amount of time taken to resolve the two types of conflict, and the number of conflicts of each type judged to be easier to resolve.

The mean resolution times taken for each type of conflict by the male and female groups are con-

TABLE 1  
MEAN RESOLUTION TIMES (IN SEC.) FOR  
APPROACH-APPROACH AND AVOIDANCE-  
AVOIDANCE CONFLICTS OF MALE  
AND FEMALE SUBJECTS

Sex	Conflict	
	Approach-Approach	Avoidance-Avoidance
Male	128.92	164.32
Female	115.08	142.00
Both	122.00	153.16

TABLE 2  
ANALYSIS OF VARIANCE OF RESOLUTION  
TIME FOR AVOIDANCE-AVOIDANCE AND  
APPROACH-APPROACH CONFLICTS

Source of variation	df	Mean square	F
Between conflicts	1	38844.06	110.28*
Between sexes	1	13086.30	3.14
Conflict X sex	1	718.26	2.04
Between Ss within groups	78	4172.52	
Conflict X Ss within groups	78	352.22	
Total	159		

\* Significant at the .01 level.

TABLE 3  
MEAN NUMBER OF APPROACH-APPROACH AND  
AVOIDANCE-AVOIDANCE CONFLICTS  
JUDGED EASIER TO RESOLVE

Sex	Conflict	
	Approach-Approach	Avoidance-Avoidance
Male	12.72	8.28
Female	12.55	8.45
Both	12.64	8.36

tained in Table 1. The results of an analysis of variance, presented in Table 2, indicate that the difference between the time taken to resolve the approach-approach and avoidance-avoidance conflicts was significant at the .01 level of confidence. The interaction between sex and the experimental condition was not significant. The data were further analyzed for possible differences in variability between sexes; at the .05 level of confidence, the hypothesis of equal variability could not be rejected ( $F = 1.03$  for approach-approach scores and 1.48 for avoidance-avoidance scores,  $df = 39$  and 39).

The mean numbers of each type of conflict judged easier to resolve by the male and female groups are contained in Table 3. An analysis of the difference in variability between sexes showed that at the .01 level of confidence the males were more variable than the females ( $F = 1.98$ ,  $df = 39$  and



39). A  $t$  test of the type suggested by Gronow (3) for use with groups differing in variability indicated that at the .05 level of confidence the null hypothesis in respect to sex differences could not be rejected ( $t = .20$ ,  $df = 39$  and  $39$ ). Assuming the null hypothesis tenable, the two sex groups were then combined to test the hypothesis that the mean number of approach-approach conflicts judged easier to resolve minus the mean number of avoidance-avoidance conflicts similarly judged was zero; the results indicated that the null hypothesis could be refuted considerably beyond the .01 level of confidence ( $t = 5.74$ ,  $df = 79$ ).

#### CONCLUSION

The results of the present experiment are in agreement with Lewin's deduction that approach-approach conflicts are more easily resolved than are avoidance-avoidance conflicts. Approach-approach conflicts required significantly less time to resolve than did avoidance-avoidance conflicts, and significantly more approach-approach conflicts were judged easier to resolve than were avoidance-avoidance conflicts. Males and females

did not appear to differ in the conflict behavior except that in judging the difficulty of conflicts, males proved to be significantly more variable than females.

#### REFERENCES

1. BARKER, R. G. An experimental study of the resolution of conflict by children. In Q. McNemar & Maude A. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942. Pp. 13-34.
2. BARKER, R. G. An experimental study of the relationship between certainty of choice and the relative valence of the alternatives. *J. Pers.*, 1946, 15, 41-52.
3. GRONOW, D. G. C. Test for the significance of the differences between means in two normal populations having unequal variances. *Biometrika*, 1951, 38, 252-256.
4. HOVLAND, C. I., & SEARS, R. R. Experiments on motor conflict. I. Types of conflict and their modes of resolution. *J. exp. Psychol.*, 1938, 23, 477-493.
5. LEWIN, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.

Received October 15, 1956.

## PERSONALITY CHANGE AS A RESULT OF AN EXPERIMENTAL CHANGE OF ENVIRONMENTAL CONDITIONS

ARNOLD S. TANNENBAUM

*Survey Research Center, University of Michigan*

THE QUESTION of the consistency of adult personality is a continuing problem in the field of personality theory. Kelly, in his longitudinal study of marital adjustment, found evidence of a remarkable degree of consistency as well as some evidence of significant changes over a 20-year period (2). One of the questions that remains is the extent to which personality change, when it occurs, can be understood in terms of the effect of specific environmental conditions. The present note is intended to report results bearing on this question from an experimental study in a large clerical organization. Other aspects of this study have been reported earlier (4, 5, 7).

Two experimental groups were created that differed in their distribution of control. In one, called the autonomous program, the amount of control exercised by lower-level employees was increased and that of the upper hierarchical levels decreased. In the other, called the hierarchical program, the amount of control exercised by the rank and file was decreased and that of the upper levels increased. The program for each group of approximately 200 female employees continued for over a year. A paper-and-pencil questionnaire

designed specifically to measure 26 personality trends to which each of the programs had some degree of relevance was administered near the beginning of the experimental period, and about a year later at the end. Results reported earlier (7) indicated that subjects' reactions to the work programs can be understood partly in terms of the relationships between their personality structures and the nature of the program in which they were located. It is of further interest, however, to investigate personality as a dependent variable; the extent to which subjects' personality trends are found to change in predictable ways as a result of their experiences in the experimental programs.

It was initially assumed that personality trends are quite stable and would be unlikely to change as a result of experimental treatment.<sup>1</sup> Nevertheless, the two programs were seen to have profound effects on the subjects' attitudes toward the company and satisfactions in it. In addition, the duration of the experiment, as well as the seriousness with which the subjects reacted to it, raised the question of its possible effects on the relevant

<sup>1</sup> For the conception of personality used, see Allport, F. H. (1).



TABLE 1

## BEFORE-AFTER CORRELATIONS AND INCIDENCE OF CHANGE

(Change in the predicted direction significant at the .05 level is indicated by +; change in a direction opposite to that predicted by -)

	Before- after ( <i>N</i> = 195) <sup>a</sup>	Autono- mous Program ( <i>N</i> = 115) <sup>a</sup>	Hierar- chical Program ( <i>N</i> = 80) <sup>a</sup>
<i>Autonomous Trends</i>			
1. assume responsibility (anticipating, preparing for, and accepting the consequences of one's activities)	.59	n.s.	n.s.
2. be creative (develop new patterns; to use one's imagination; to think up new and original ways of doing things)	.56	n.s.	n.s.
3. increase the variety (number of types) of satisfactions in any given situation (to enjoy as many different things as possible in the same situation, e.g., to "mix business with pleasure")	.41	n.s.	n.s.
4. understand and explain the reasons for one's behavior and opinions (to understand and explain one's feelings)	.48	n.s.	n.s.
5. help others	.59	n.s.	n.s.
6. understand others' viewpoints and feelings (to know why people feel or act as they do; to understand the reasons for others' behavior)	.60	n.s.	+
7. be independent of control figures (to be free of direction and control of persons in authority)	.54	n.s.	+
8. express one's considered judgments (to think through things carefully and state one's opinions)	.78	n.s.	+
9. base one's actions on one's own critical judgments and evaluations (to carefully evaluate and decide things for oneself)	.38	-	+
10. act on the same level as others (to treat and be treated by others as an equal)	.52	n.s.	n.s.
11. understand relations between and reasons for things (to understand the basis for things; to figure things out; to know "why")	.72	-	+
12. do one's own thinking (not be suggestible; to decide things for oneself; to draw one's own conclusions)	.40	n.s.	+
13. take the initiative with others (to start things; to organize things)	.42	n.s.	-
<i>Hierarchical Trends</i>			
14. avoid emotional involvements (to avoid the expression and reception of emotional acceptance; to keep emotionally distant from others; to avoid being "warm" in relationships with others; to avoid affection)	.65	n.s.	n.s.
15. conform to the wishes of control figures (to please persons in authority)	.60	n.s.	n.s.
16. adhere rigidly to specific rules or directions (to follow rules to the letter; to be given directions for doing things)	.58	n.s.	n.s.
17. act inferior to control figures (to be humble in relation to persons in authority)	.60	+	n.s.
18. avoid committing oneself (to keep one's opinions to oneself; to keep from getting oneself "out on a limb"; to keep people from knowing where one stands on controversial questions)	.40	n.s.	n.s.
19. keep one's roles (behavior categories, relationships) at any given time distinct from other roles (to keep various parts of one's life separated)	.40	n.s.	+
20. admire or respect control figures or symbols (to look up to persons in, or symbols of, authority)	.68	n.s.	n.s.
21. be superior to others (to be better than others; to be more outstanding than others)	.69	+	-
22. be efficient (not waste time or effort)	.58	+	-
23. be submissive to control figures (to submit to dominant leaders; to obey strong and forceful authority figures)	.45	n.s.	+
24. show self-discipline (to show self control; to keep one's behavior disciplined)	.42	+	-
25. depend on others (to receive direction from others)	.48	n.s.	n.s.
26. obey rules and follow directions	.58	n.s.	n.s.

<sup>a</sup> In some cases the *N*'s are slightly less than this value due to a very small proportion of NA's—never more than 4%.

personality variables. A number of hypotheses regarding change in personality trends were therefore formulated on the basis of the assumption that environmental conditions may increase or decrease the characteristic energetic (homeostatic) level of trends on the basis of a simple reward-and-punishment principle: characteristics that are given opportunities for expression tend to increase

in potency; trends that are given only minimal opportunities for expression tend to decrease.

The 26 trends measured by the personality questionnaire were divided into two groups of thirteen: "autonomous trends," which were likely to be given opportunities for expression in the autonomous program (and not in the hierarchical), and "hierarchical trends," which were likely to be



given opportunities for expression in the hierarchical program (and not in the autonomous). It was hypothesized that (a) in the autonomous program, autonomous trends should tend to increase in potency and hierarchical trends tend to decrease; (b) in the hierarchical program, hierarchical trends should tend to increase in potency and autonomous trends tend to decrease.

Care was taken in framing the questions used to measure the personality trends to make them as general as possible and distinct from the company situation. Responses to these questions are therefore not likely to reflect the subjects' changing attitudes toward the experimental programs or toward the company. Five questions were employed to measure each trend. For example, one of the questions for the trend, "trying to be independent of control figures," was "How much do you believe that people generally require someone to tell them what to do?" For the trend, "trying to understand others' viewpoints and feelings," we asked, "How much does it bother you when you do not understand the viewpoints and feelings of people you associate with?" For the trend, "trying to keep one's role at any given time distinct from other roles," we asked, "Generally, how good an idea do you think it is to choose very close friends from among persons you work with?" For the trend, "trying to do one's own thinking," we asked "How hard do you find it to disagree with others even in your own thinking?" For the trend, "trying to act inferior to control figures," we asked, "Do you think that people who have reached high office in government have a right to feel that they are better than most people?" For the trend, "trying to be submissive to control figures," we asked, "Do you think that for a truly happy marriage that in major issues the husband should tell the wife what to do?"

## RESULTS

Table 1 shows the before-after correlations (test-retest reliabilities) for each of the trends, and the direction of changes significant at the .05 level or less. A plus sign indicates that the change is in the direction predicted; a minus sign, a change in the opposite direction. Although all of the predictions are directional, two-tailed rather than one-tailed tests are employed to permit ascribing levels of confidence to changes contrary to hypothesis.<sup>2</sup>

Twelve changes in the predicted direction prove significant at the .05 level of confidence. Six changes, significant at the .05 level occur in a direction opposite to that predicted.<sup>3</sup> The combined

results cannot be assessed statistically since the trend measures are not independent, but it seems reasonable to conclude that some change has occurred and that it is predominantly in the predicted direction.

The preponderance of changes in the direction of decreasing trend potency is hard to explain. Most of the "correct" changes involve the decrease of autonomous trends in the hierarchical program or the decrease of hierarchical trends in the autonomous program. Likewise, five of the six "incorrect" changes involve the decrease of autonomous trends in the autonomous program or hierarchical trends in the hierarchical program. No single trend is found to change significantly in the predicted direction in both programs.

Although somewhat ambiguous, the data seem to indicate that measurable change can be effected by a persisting change in environmental conditions. Furthermore, the change seems partly explicable in terms of the movement of personality toward equilibrium with its environment. A type of organism-environment reciprocity is suggested in which the organism selects, reacts, and adjusts to its environment, so as to optimize personality "closure" while at the same time modifying itself within limits so as to further increase trend expression. This integral character of inner needs and outer conditions, as a general psychological principle is consistent with the position taken by Murphy: "A personality is a structured organism-environment field, each aspect of which stands in dynamic relation to each other aspect. There is organization within the organism and organization within the environment, but it is the cross-organization of the two that is investigated in personality research" (6, p. 8).

## REFERENCES

1. ALLPORT, F. H. Teleonomic description in the study of personality. *Charact. Pers.*, 1937, 5, 202-214.
2. KELLEY, E. L. Consistency of the adult personality. *Amer. Psychologist*, 1955, 10, 659-681.
3. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
4. MORSE, NANCY, & REIMER, E. The experimental change of a major organizational variable. *J. abnorm. soc. Psychol.*, 1956, 52, 120-129.
5. MORSE, NANCY, REIMER, E., & TANNENBAUM, A. S. Regulation and control in hierarchical organizations. *J. soc. Issues*, 1951, 7, (3), 41-48.
6. MURPHY, G. *Personality, a biosocial approach to origins and structure*. New York: Harper, 1947.
7. TANNENBAUM, A. S., & ALLPORT, F. H. Personality structure and group structure: an interpretative study of their relationship through an event structure hypothesis. *J. abnorm. soc. Psychol.*, 1956, 53, 272-280.

Received February 18, 1957.

In this case, eight changes prove "correct" and four "incorrect."

<sup>2</sup> The following statistical formula, designed specifically to test the significance of change, was employed:  $t = (A - D - 1/\sqrt{A + D}) / (3, p. 207)$ .

<sup>3</sup> The pattern of results remains essentially the same when evaluated in terms of the .01 level of confidence.